Semiparametric Estimation of Locally Stationary Diffusion Models*

Bonsoo Koo†
London School of Economics

Oliver Linton‡
London School of Economics

The Suntory Centre
Suntory and Toyota International Centres for
Economics and Related Disciplines
London School of Economics and Political Science
Houghton Street
London WC2A 2AE
Tel: 020 7955 6679

Discussion paper
No. EM/2010/551
August 2010

* We would like to thank Yacine Aït-Sahalia, Dennis Kristensen and Peter Phillips for helpful comments.

† Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom; e-mail: b.koo@lse.ac.uk

‡ Department of Economics, London School of Economics, Houghton Street, London WC2A 2AE, United Kingdom. E-mail address: o.linton@lse.ac.uk. Thanks to the European Research Council for financial support. This paper was partly written while I was a Universidad Carlos III de Madrid-Banco Santander Chair of Excellence, and I thank them for financial support.

Abstract

This paper proposes a class of locally stationary diffusion processes. The model has a time varying but locally linear drift and a volatility coefficient that is allowed to vary over time and space. We propose estimators of all the unknown quantities based on long span data. Our estimation method makes use of the local stationarity. We establish asymptotic theory for the proposed estimators as the time span increases. We apply this method to the real financial data to illustrate the validity of our model. Finally, we present a simulation study to provide the finite-sample performance of the proposed estimators.

Key words: diffusion processes, local stationarity, term structure dynamics, density matching, option pricing.

Journal of Economic Literature Classification: C14, C32

1 Introduction

The theory of asset pricing has been one of the fastest growing fields of study over the past decades. Since the seminal papers written by Black and Scholes (1973) and Merton (1973), the diffusion process has lain at the heart of modeling the dynamics of economic variables, including the term structure of interest rates. This defines X_t as the strong solution to

$$dX_t = \mu_t dt + \sigma_t dW_t, \tag{1}$$

where $\{W_t : t \geq 0\}$ is a standard Brownian motion defined on the filtered probability space $(\Omega, \mathcal{F}^W, (\mathbb{F}_t^W), P)$, $\mathcal{F}^W$ is a σ-algebra, $\mathbb{F}_t^W$ is a filtration. Here, μ_t and σ_t^2 are commonly referred to as the conditional drift or instantaneous return function, and the conditional diffusion or volatility function of the process respectively. They can depend on X_t and time.

In line with the remarkable progress in asset pricing theory, the estimation methodologies concerning continuous-time stochastic processes have improved immensely over the most recent decade. Because one of the main goals of financial econometrics is to investigate the expected returns and volatilities of the underlying dynamics of economic variables such as stocks, interest rates, exchange rates, and their derivatives, the econometric treatment of estimating the above two functions of interest has advanced quite significantly and become more sophisticated.

In spite of the progress on various fronts of econometric theory on continuous-time stochastic processes, quite a few challenges still remain to be addressed. For instance, identification, estimation and studies of the asymptotic properties of the continuous-time processes have turned out to be quite demanding, mainly because we only have discretely sampled observations drawn from processes whose dynamics are continuous in time. Moreover, while the diffusion process in (1) is quite extensively used, asset pricing theory doesn't narrow down the number of possible specifications for the drift and volatility terms, let alone pin down their exact forms. For example, an array of different specifications have been proposed for the term structure dynamics (see Aït-Sahalia (1996)).

As a result, many stochastic models have been chosen simply due to mathematical manipulability and simplicity of statistical inferences. One salient example could be the assumption of stationarity, i.e., the assumption of the existence of a time invariant stationary distribution for X_t. Indeed, most of the financial econometrics theories depend on the assumption of stationarity of the observed process. Undoubtedly, this is because a stationarity assumption provides a powerful device for identification and estimation of the underlying continuous-time data generating process. More specifically, there are conspicuous benefits attributed to this assumption. Among them, most importantly, it enables us to avoid the serious identification issue known as the aliasing problem, since cross-restrictions can be imposed. It is worth mentioning that in general, there is no one-to-one correspondence (bijection) between the parameters or functionals of the continuous-time model and its corresponding discrete time model, see Phillips (1973). Moreover, under suitable conditions, a strict stationarity assumption guarantees that the distribution of the diffusion process is completely characterized by two functionals of our concern. Aït-Sahalia (1996) and Jiang and Knight (1997) used this property to substantiate their arguments. Another compelling reason for the stationarity assumption is because well-established asymptotic results are readily available. Therefore, it makes establishment of estimation and inference procedures much simpler.

Nevertheless, there are a plethora of cases in which an assumption of stationarity is unrealistic and unjustifiable. For example, irregularity of the trade at the beginning or the end of financial markets, volatility clustering, and ruptures arising from shocks or structural changes are among those (see Campbell, Lo and Mackinlay (1997)). In truth, nonstationary properties of financial data are often found in many problems of interest in economics and finance. In addition, time series in economic and financial markets have an inherently dynamic and time-varying nature. Therefore, the stationarity assumption is unlikely to hold in many cases. Consequently, it would be desirable that nonstationarity could be allowed for. Not only does this relaxation enable us to fit the data better, but may also allow us to augment the model built upon the assumption of stationarity to a more general model with nonstationary characteristics.

This paper proposes a semiparametric model that generalizes the usual time-inhomogeneous diffusion processes along the line of an important class of nonstationarity, *local stationarity,* Dahlhaus (1996). This type of nonstationarity is relatively easy to deal with since, in a neighborhood of a chosen time point the process behaves like a stationary diffusion process. We use this structure to devise an estimation method. More specifically, our approach is based on the density matching method of Aït-Sahalia (1996) suitably generalized to this nonstationary framework. We will discuss relevant papers in the Section 2 in more detail.

We establish the asymptotic properties of our estimators under the long span assumption only, namely we do not require that the time between observations goes to zero, rather we assume that it is fixed but the horizon under consideration expands. This seems like an appropriate framework for certain types of data. We show that our proposed estimators of the drift and volatility of processes are consistent and asymptotically normal. In addition, we obtain the uniform rate of convergence of the estimators of the drift and the volatility functions. Finally, we present a simulation study and an application to weekly interest rate data to illustrate the finite sample properties of the proposed estimators of the drift and diffusion.

The remainder of this paper is organized as follows. Section 2 briefly reviews related studies in the literature. Section 3 introduces the model and the related framework. Section 4 suggests estimators of the drift and volatility functions. Section 5 develops several asymptotic theories in relation to our proposed estimators. Also, an real application of our methodology is given in Section 6. Simulation results are discussed in section 7. Section 8 concludes. All proofs and our application and simulation results are found in Appendix.

Throughout this paper, the following notations are used. The integral $\int$ is taken over $(-\infty, \infty)$ unless specified otherwise. $||\cdot||$ denotes any norm over the relevant space. Let g be any function. $||g||_\infty = \sup_x |g(x)|$, $||g||_2 = \left(\int |g(x)|^2 dx\right)^{1/2}$ and $||g||_2^2 = \left(\int |g(x)|^2 dx\right)$. $\mathbb{C}_2(b)$ denotes the space of twice continuously differentiable real valued functions with first and second partial derivatives of all of their arguments bounded by b and $||g||_\infty < b$. In particular, in the real valued function of two variables, say $g(u, x)$, u denotes a time index. If there is no confusion, terms like $g(u, x)$ will often be written as $g_u(x)$ to emphasize our focus on the neighborhood of a time point u. Meanwhile, a subscript, 0 denotes the true value or functional form of the corresponding parameters or functions. $g^{(m)}(u)$ denotes the mth derivative with respect to u and $g^{(m)}(u,x)$ denotes the mth derivative with respect to x whereas $\dot{g}(u,x)$ and $\ddot{g}(u,x)$ denote the first and second derivative with respect to u. All convergences are considered when $T \to \infty$.

2 Related Literature

2.1 Non- and semi-parametric estimation methods

Given that an array of specifications can be compatible with asset pricing theory with no arbitrage condition, a fully parametric model is more likely to be subject to misspecification. To prevent possible misspecification problems, a larger family of parametric models with some degree of flexibility is required. Nonparametric and semiparametric estimation methods are suitable to ensure a high flexibility since any restriction on functional forms of objects of interest is kept to a minimum. In addition, given that kernel estimators are data-driven estimation methods, the rapid development of capacity of computing and the availability of immense amount of financial data make this method more attractive.

Consequently, these methods have become more popular and played an increasingly important role in estimation and inference of the diffusion processes. For instance, fully nonparametric estimation methods are used in Jiang and Knight (1997), Stanton (1997), and Bandi and Phillips (2003). Meanwhile, Aït-Sahalia (1996) and Kristensen (2004) considered semiparametric methods (see Kristensen (2004) for an overview). Among those, we focus on semiparametric methods along the line of density matching.

Aït-Sahalia (1996) and Jiang and Knight (1997) used density matching to estimate both functionals semiparametrically and nonparametrically respectively. Density matching implies the one to one mapping between the drift and diffusion functions and the marginal and transitional densities. More specifically, Aït-Sahalia (1996) starts with realization of the equivalence between (μ, σ^2) and marginal and transitional densities of the process. He proposed the asymptotic theory with respect to the nonparametric estimator of diffusion function under the context of stationarity and long span asymptotic. However, to make use of this methodology, strong stationarity of the underlying process is required since it utilizes the Kolmogorov forward equation which is only valid under the assumption of stationarity. Later, Kristensen (2004) extended this work further with respect to more general classes of semiparametric diffusion models. These papers, however, are dependent upon the assumption of the stationarity in the sampled diffusion process. Without it, the results are likely to break down.

2.2 Estimation of Nonstationary diffusion processes

It is worth mentioning that there are too many classes of nonstationary processes to construct a valid econometric argument without restricting attention to a specific type of nonstationarity. One way of dealing with nonstationarity is to restrict ourselves to a class of processes, for which we can specify how they are different from stationarity without losing much generality.

There are a few attempts on the class of nonstationary diffusion processes. Bandi and Phillips (2003) lead one strand of literature in which the recurrence assumption plays an essential role in addressing one class of nonstationarity processes. This recurrence assumption implies the continuous trajectory of the process visits any level in its range an infinite number of times over time. It guarantees enough visits in the neighborhood of a certain points in the range of the process that an infinite number of differences can be averaged asymptotically, which ensures consistent estimators can be acquired in the limit. Under this scheme, consistency of the estimators of both functional forms of the drift and diffusion could be obtained under the joint implementation of infill and long span asymptotics. More specifically, these asymptotics require that the distance between observations Δ goes to zero in the limit (infill) while the time span T goes to infinity (long span). The basic idea in their paper is that long span asymptotics enables us to make use of recurrence properties of the process whereas infill asymptotics ensures infinitesimal characteristics of both the drift and diffusion. Bandi and Nguyen (2003) extended this idea to the estimation of a Jump-diffusion model.

In another strand of literature on nonstationary processes, Fan *et al* (2003) introduce a family of time-dependent diffusion processes to allow for time-changing nature of economic variables. They assumed that the two functionals of interest are time-varying and smoothly evolving. Their idea is based on local smoothness in the time domain. They make use of infill asymptotics in their analysis.

Our approach is mainly different from Bandi and Phillips (2003) in that whereas Bandi and Phillips (2003) make use of infill and long-span, we look closely at the limiting behaviour of the time span T, i.e. long-span only, to establish the consistency of the proposed estimators and obtain their limit distributions thanks to the assumption of local smoothness of functions of both the drift and the diffusion. In addition, our method is different from Fan *et al* (2003) in that we make use of the concept of density matching to derive our estimators and establish asymptotic theories of them. To the best of our knowledge, this is the first *local stationarity* treatment of establishing asymptotics of both the drift and the diffusion of time-inhomogeneous diffusion processes.

2.3 Local stationarity

Although econometric treatment of continuous-time nonstationary processes has a relatively short history, the analyses of the class of time-varying processes in discrete time have evolved quite extensively both in the frequency and time domain. However, since we approach the diffusion processes in the time-domain, we will focus our attention on time-domain analysis. In particular, we limit our focus to a family of evolutionary or locally stationary processes in continuous time. Local stationarity, one kind of local time concept, will serve as an important identifying condition. Heuristically, local stationarity implies that a process behaves in a stationary manner only in a neighbourhood of a given time point but is nonstationary over the whole horizon.

The early treatment of time varying coefficients of multiple regression models in discrete time can be found in Robinson (1989). In his paper, the asymptotics is fully developed showing asymptotic normality and consistency of nonparametric estimators for time varying parameters. In his model, however, parameters of interest depend solely on time and the main results are based on stationary state variables X_t. As a result, the nonstationarity only comes from time varying coefficients.

The statistical treatment of this type of nonstationarity has undergone significant breakthroughs following a series of papers by Dahlhaus (1996, 1997) and Dahlhaus and Polonik (2009). These papers provided a more rigorous definition and treatment of locally stationary processes. Moreover, whereas most early works of these types were mainly concerned with the analysis in the frequency domain, his papers were mainly interested in parametric inference for nonstationary models defined purely in the time domain. Given that many models in economics and finance are designed in the time domain framework, his theory could be the most relevant to our efforts to capture the asymptotic behaviour of our estimators in the diffusion processes as a solution to SDE widely used in the literature.

However, there are several limitations that are worth mentioning. Dahlhaus' theories were not general enough to be applicable to our model directly. He proposed a statistical inference procedure in univariate autoregression model. He relied on a theory of evolutionary spectra and therefore, his theories are applicable only to a class of simple AR models. In addition, he took a fully parametric approach and assumed specific functional forms for time-varying processes. Later, Kim (1999) extended it to multiple regression models using nonparametric kernel methods. Orbe *et al* (2000) wrote a series of papers on time varying estimation as well. Quite recently, Dahlhaus and Subba Rao (2006) and Fryzlewicz *et al* (2008) studied a time-varying ARCH processes via local Maximum Likelihood method and local Least squares method respectively.

Since we are concerned with continuous-time diffusion processes as in (1), the results of the above papers should be appropriately adjusted since they are based on discrete time processes. Whereas some important concepts could still go through analogously, there are complications arising from the fact that we are dealing with the continuous-time processes based on discretely sampled observations since, in reality, the continuous account of observations are unattainable.

3 Model

The underlying data generating process of interest $\{X_{t,T}; t, T \geq 0\}$ is a time-inhomogeneous Itô diffusion process represented by the following traingular array of stochastic differential equations

$$dX_{t,T} = \mu(t/T, X_{t,T})dt + \sigma(t/T, X_{t,T})dW_t, \qquad X_{0,T} = x_0, \tag{2}$$

where $\mu(u, x) = \beta(u)(\alpha(u) - x)$ and $\mu(\cdot)$, $\sigma^2(\cdot)$ and $\{W_t\}$ are defined as in (1). x_0 is a given random variable. $\{X_{t,T}\}$ defined above is a one-dimensional process and has a domain $\mathbf{I} = [l, r]$ where $-\infty \leq l < r \leq \infty$. Note that (2) defines a triangular array of observations $X_{t,T}$, but we shall on occasions use the simpler notation that dispenses with the T subscript.

We will approximate the above process by a family of stationary processes indexed by u,

$$d\tilde{X}_{u,t} = \mu(u, \tilde{X}_{u,t})dt + \sigma(u, \tilde{X}_{u,t})dW_t, \qquad \tilde{X}_{u,0} = x_{u,0} \tag{3}$$

where, in the vicinity of each fixed time point $u \in [0,1]$, (C1) both drift and diffusion functions have at least twice bounded and continuous derivatives; (C2) $\sigma^2(u,x) > 0$ for $\forall x \in \mathbf{I}$; (C3) for $\forall x \in \mathbf{I}$, $\exists \varepsilon > 0$ such that $\displaystyle\int_{x-\varepsilon}^{x+\varepsilon} \frac{|\mu(u,y)|dy}{\sigma^2(u,y)} < \infty$. (C4) the scale function defined as $S(u,x) = \displaystyle\int_c^x s(u,\xi)d\xi$ where $s(u,\xi) = \exp\left\{-2\displaystyle\int_c^{\xi} \frac{\mu(u,\zeta)d\zeta}{\sigma^2(u,\zeta)}\right\}$ for $x \in \mathbf{I}$, satisfies the following: $\int_l^c s(u,\xi)d\xi \to -\infty$, $\int_c^r s(u,\xi)d\xi \to \infty$ for some fixed number $c \in \mathbf{I}$. (C5) the speed measure defined as $\int_l^r \left[s(u,x)\sigma^2(u,x)\right]^{-1} dx$ is bounded. It is worth noting that under C1-C5, for each time point $u \in [0,1]$, $\{\tilde{X}_{u,t}; t \geq 0\}$ are strictly stationary and weakly dependent with a stationary density $f_0(u,x)$. (See Karatzas and Shreve (2000).) Moreover, $\{\tilde{X}_{u,t}; t \geq 0\}$ are β-mixing. (See Chen *et al.* (2008).)

Equations (2) and (3) can be equivalently written as

$$X_{t,T} = X_0 + \int_0^t \mu\left(\frac{s}{T}, X_{s,T}\right) ds + \int_0^t \sigma\left(\frac{s}{T}, X_{s,T}\right) dW_s, \tag{4}$$

$$\tilde{X}_{u,t} = \tilde{X}_0 + \int_0^t \mu\left(u, \tilde{X}_{u,s}\right) ds + \int_0^t \sigma\left(u, \tilde{X}_{u,s}\right) dW_s. \tag{5}$$

Also, we can think of $\{\tilde{X}_{t/T,t}\}$ for t/T in a neighborhood of u as follows.

$$\tilde{X}_{t/T,t} = \tilde{X}_0 + \int_0^t \mu\left(\frac{t}{T}, \tilde{X}_{t/T,s}\right) ds + \int_0^t \sigma\left(\frac{t}{T}, \tilde{X}_{t/T,s}\right) dW_s. \tag{6}$$

For t/T in a neighborhood of u, we can approximate $X_{t,T}$ by $\tilde{X}_{u,t}$ in the sense that local moments calculated under the distribution of $X_{t,T}$ are close to the same moments calculated under the distribution of $\tilde{X}_{u,t}$. To this end, we define a locally stationary stochastic process as follows.

Definition 1 *The stochastic process $\{X_{t,T}\}$ represented by (2) is called locally stationary if there exists a stochastic process $\{\tilde{X}_{u,t}\}$ represented by (3), which is the time-homogeneous Itô diffusion process associated with $\{X_{t,T}\}$ at a given time point u for $\forall u \in (0,1)$ such that*

$$\Pr\left\{\max_{1\leq t\leq T}\left|X_{t,T}(\omega) - \tilde{X}_{t/T,t}(\omega)\right| \leq D_T T^{-1/2}\right\} = 1 \tag{7}$$

for all T, where $\{D_T\}$ is a well-defined positive process satisfying for some $\eta > 0$,

$$E\left(|D_T|^{4+\eta}\right) < \infty.$$

If the underlying process is stationary locally in time, $X_{t,T}$ and $\tilde{X}_{t/T,t}$ should be close on a time segment around a given time point where the assumption of stationarity holds. In our definition, the convergence rate between $X_{t,T}$ and $\tilde{X}_{t/T,t}$ is given as $O_p\left(T^{-1/2}\right)$ and it is worth noting that this rate is not exact but rather conservative. We show this rate under our smoothness assumption in Appendix A even though the exact convergence rate of the above condition (7) is hard to obtain mainly because those stochastic processes are not fully parameterized. However, heuristically, it seems plausible that if $X_{t,T}$ is locally stationary, $X_{t,T}$ and $\tilde{X}_{t/T,t}$ should be close and the degree of approximation should depend on the rescaling factor T. In this regard, we assume the absolute value of the differences between two processes can be uniformly bounded by $D_T T^{-1/2}$ where the moment of $|D_T|^{4+\eta}$ is uniformly bounded. This moment condition is required mainly for asymptotic normality. Comparing (5) and (6), due to differentiability and Talyor expansion, the degree of approximation between $\tilde{X}_{u,t}$ and $\tilde{X}_{t/T,t}$ should rely on the difference between t/T and u. These will be explored shortly and be used frequently in Section 5.

Throughout the paper, our attention focuses on the locally stationary diffusion processes.

Assumption 1 *We observe realizations $\{X_{t,T}\}_{t=1}^{T}$ from a locally stationary process represented by (2).*

It is worth noting several features of our model. Firstly, the specification (2) is analogous to the local fit of the usual nonparametric estimation methods especially in the nonparametric curve fitting literature. This ensures the amount of *local* information increases appropriately when T increases. It also highlights that we are concerned with the vicinity of a given time point, u. Note that this is not the case in which the sampling of observations occurs on (0,1) nor the case in which the interval between two contiguous time points goes to zero (see Robinson (1989) and Dahlhaus (1997)). This estimation method is built upon the assumption of local stationarity. In more detail, let $g_t(x)$ and $f_u(x)$ be the density functions of $\{X_{t,T}\}$ and $\{\tilde{X}_{u,t}\}$ respectively. We assume that for each given $u \in (0,1)$, $\{\tilde{X}_{u,t}\}$ is the stationary diffusion process which has the same finite-dimensional distributions as $\{X_{t,T}\}$ in the vicinity of a corresponding time point u and hence, there exists a mapping from u to a density function, i.e. $u \longmapsto f_u(x)$ where $f_u(x)$ is the density function associated with $g_t(x)$ at a given time point u. Each time point has a possibly different distribution and therefore there exists a family of density functions $\{f_u : u \in (0,1)\}$. For a fixed time point u, we can estimate a density function from which the drift and the diffusion functions can be derived using stationarity properties. That is, we imagine we choose a tiny region around u and treat the process of interest as stationary on that region with the drift and the diffusion. To this end, we assume the drift and the diffusion don't change much in the vicinity of a time point of interest since the smoothness of μ and σ in (2) ensures that $\{X_{t,T}; t \geq 0\}$ is locally stationary at least asymptotically. In other words, both μ and σ are smooth functions which allow for a certain level of differentiability.

Next, our parameterized drift term takes the form

$$\mu(u,x;\theta(u)) = \beta(u)(\alpha(u) - x), \tag{8}$$

where $\alpha(\cdot)$, $\beta(\cdot)$ are unknown functions. That is to say, our focus in this paper is mainly on the diffusion process in which the drift term is parameterized locally in time up to an unknown parameter function vector $\theta = [\alpha, \beta]^\top \in \Theta$ where Θ is a parameter function space and the volatility term with no restriction. Such a parameterization of the drift function ensures that we can identify both the drift and the diffusion functions based on discretely sampled observations. This is due to the well-known fact that identifying both the drift and diffusion functions is unattainable without sufficient restrictions. Also, it allows us to adopt a local least squares estimation method for the drift. Especially in this paper, along the line with Aït-Sahalia (1996), we parameterize the drift in a way that the drift captures several features of movement of interest rates such as a mean-reverting property as seen in the following Figure 1.

FIGURE 1 ABOUT HERE

Our parameterization is coherent with various spot interest rates models widely used in the literature. Throughout the paper, we will focus on this form for the drift term. However, we think our approach can be extended to a more general specification as in Fan *et al* (2003) with a local quasi-maximum likelihood estimation under additional regularity conditions as in Kristensen (2004).

Last but not least, note that our approach is based on discretely sampled observations. Recall that we consider a triangular array of random variables $\{X_{t,T}\}$, $t = 0, 1, \ldots, T + \Delta$, $T = 1, 2, \ldots$. Suppose we conduct the following sampling. We observe the process of $X_{t,T}$ at time $t = 0, 1, \ldots, T$ over the time span $[0, T]$ while the underlying process evolves continuously in time. We assume we can't observe the continuous evolvement between two sampling points. In particular, the data is sampled at equally spaced time, say $\Delta \equiv 1$. Most importantly, the distance between two sampling points, Δ, won't shrink but remains fixed. We will be looking closely at the limiting behaviour of our estimators of both functions of interest as the sampling period T tends to infinity while Δ is constant. In the following, X_t represents $X_{t,T}$ unless otherwise specified for the simplicity of notation.

Under this framework, the following propositions ensure the validity of our econometric treatment.

Proposition 1 *Suppose Assumption 1 holds. Then, due to Definition 1,*

$$\left| X_t - \tilde{X}_{u,t} \right| = O_p\left(\left| \frac{t}{T} - u \right| + T^{-1/2} \right) \tag{9}$$

The proposed estimators and model are defined with the data, $\{X_t\}_{t=1}^T$ while our asymptotic analysis is based on the collection of locally stationary processes with $\{\tilde{X}_{u,t}; t \geq 0\}$. Proposition 1 allows us to approximate $\{X_t\}$ by a group of $\{\tilde{X}_{u,t}\}$ as $T \to \infty$ since it implies that in the vicinity of a time point u, we can replace X_t by $\tilde{X}_{u,t}$ for our asymptotic analysis since the local moments from the distribution of X_t can be approximated by those from the distribution of $\tilde{X}_{u,t}$.

Proposition 2 *Let $g_t(x)$ and $f_u(x)$ be the density functions of $\{X_t\}$ and $\{\tilde{X}_{u,t}\}$ respectively. Suppose Assumption 1 holds. Then, due to Definition 1, density functions $g_t(x)$ and $f_u(x)$ are twice boundedly continuously differentiable and have the following relationship.*

$$\|g_t(x) - f_u(x)\|_\infty = \left|\frac{t}{T} - u\right| \left\|\dot{f}_u(x)\right\|_\infty + \frac{1}{2}\left|\frac{t}{T} - u\right|^2 \left\|\ddot{f}_u(x)\right\|_\infty + o\left(\left|\frac{t}{T} - u\right|^2 + T^{-2/5}\right). \quad (10)$$

Let $g_t(x)$ and $f_u(x)$ be the density functions of $\{X_t\}$ and $\{\tilde{X}_{u,t}\}$ respectively. Under Assumption 1 and Definition 1, there exists a mapping from u to a corresponding density function associated with $g_t(x)$ at a given time point u, i.e. $\mathfrak{M} : u \longmapsto f_u(x)$. Since $\{X_t\}$ and $\{\tilde{X}_{u,t}\}$ are completely characterized by μ and σ or the transition density and the marginal density, from Definition 1, density functions $g_t(x)$ and $f_u(x)$ associated with X_t and $\tilde{X}_{u,t}$ respectively are closely related shown in Proposition 2.

4 Semiparametric Estimation

4.1 Preliminaries

4.1.1 Local Stationarity and Nonparametric Estimation

The difficulty in dealing with nonstationary processes lies in establishing the desirable asymptotic theory. Since each observation doesn't have any meaningful information on another one, letting the time span T go to infinity doesn't suffice. One way of avoiding this difficulty is to localize in time and assume that the processes of interest are stationary in the vicinity of a given time point even though it is not stationary overall.

Also, it is shown that the local stationarity assumption is milder than stationarity. All stationary processes are locally stationary but locally stationary processes do not have to be stationary. More specifically, a process is locally stationary if the process is smoothly time varying.

It is worth mentioning that in the limiting case, by construction, kernel methods focus on local properties of the process of interest. Recall that a locally stationary process behaves like a stationary process in the neighborhood of a given time point. Therefore, well established results for stationary processes can be utilized in deriving the asymptotics of the kernel estimates. The structure of the nonparametric estimates makes the derivation of the limiting theory relatively uncomplicated since the statistical tools for stationarity can be readily used in deriving the asymptotics of the proposed estimators. Specifically, under the assumption of local stationarity, kernel methods allow us to make use of Kolmogorov forward/backward equation to obtain two functionals of interest in the diffusion processes.

4.1.2 Density Matching

Density matching approach is based on realization of the equivalence between the drift and the volatility functions in one direction and marginal and transitional densities in the other. Whereas the former can not be estimated directly, the latter can be estimated directly from the data. Therefore, estimating the densities first and then finding two plausible functions of interest which correspond to the obtained densities would be a natural way to go about analyzing the diffusion processes. The Kolmogorov forward/backward equations play a crucial role in estimation of both functionals of interest under the assumption of stationarity. Density matching utilizes these equations to identify both the drift and the diffusion. In our paper, we will make an analogous argument under the context of local stationarity. The Kolmogorov backward equation is used to identify the drift term. The Kolmogorov forward equation can be generally used for computing the probability densities of stochastic differential equations. Thanks to the Kolmogorov forward equation, the density can be expressed in terms of the drift and diffusion we would like to estimate. Therefore, we can express the drift (or diffusion) term as a function of marginal density and diffusion (or drift) by inverting the equation. In our paper, due to the assumption of local stationarity and a parameterization of the drift term, we can generalize the estimation method used in Aït-Sahalia to the time-inhomogeneous diffusion processes which behave in a locally stationary manner. We replace the density function with the local analogue version. Most of the results in Aït-Sahalia (1996) can be carried over to the generality in this paper with slight modification.

In this paper, since we are concerned with time-dependent diffusion processes under the scheme of local stationarity, we consider the following argument to provide the asymptotic justification for our estimators. We localize the process in the time domain so that the above kernel estimation method enables us to restrict our attention to the range where the process behaves in a stationary manner. Against this backdrop, over the range where the process is stationary, we can generalize the methodology in Aït-Sahalia (1996) and Kristensen (2004) with slight modification. More specifically, in the vicinity of a given time point u, the process is assumed to be strict stationary. As a result, we can treat the time varying coefficient roughly constant over the suitable range $[u-h, u+h]$ around a given time point u. It is worth emphasizing that thanks to Proposition 1, we focus on (3) rather than (2) directly and therefore most important results under the assumption of stationarity can be carried over to our pursuit for estimators of two functionals in locally stationary time-dependent diffusion process. Throughout this section, we are mainly concerned with the vicinity of a certain time point u to set out our argument.

Under the assumption of local stationarity and Proposition 2, we turn to a time-localized kernel estimator to estimate the distribution around a fixed time u. In this regard, a plausible estimator for the density function can be

$$\hat{f}(u, x) = \frac{1}{Th_1h_2} \sum_{t=1}^{T} K\left(\frac{u - t/T}{h_1}\right) K\left(\frac{x - X_t}{h_2}\right),$$

where K is a real-valued kernel function concentrated around the origin and h_i, $i \in \{1, 2\}$ is a bandwidth parameter. In particular, h_1 is closely related to the size of the neighborhood of the rescaled grid point.

4.2 Estimators

4.2.1 The Drift

For the estimator of the drift function, we propose the following procedure.[1] It is worth noting that our estimation methods make use of local stationarity properties of the underlying process and the asymptotic properties of the above estimator will be analyzed based on the local approximation by the stationary process $\{\tilde{X}_{u,t}\}$ around the given time point u. Theoretically, since the drift term is parameterized as in (8), its discretized regression specification associated with the stationary process $\{\tilde{X}_{u,t}\}$ can be expressed as

$$E[\tilde{Y}_{u,t}|\tilde{X}_{u,t}] = a(u) + b(u)\tilde{X}_{u,t} \tag{11}$$

where $\tilde{Y}_{u,t} = \tilde{X}_{u,t+\Delta} - \tilde{X}_{u,t}$. In this regard, we use the local regression method to obtain the parameterized drift. With the suitably chosen bandwidth, kernel estimation method allows us to obtain locally weighted least square estimators for $a(u)$ and $b(u)$. Let the kernel function used for the estimation of the drift term be

$$K_{ut} = K\left(\frac{u - t/T}{h}\right).$$

Given the specification (11), we can estimate $[a(u), b(u)]^\top$ by minimizing the following objective function

$$\sum_{t=1}^{T}(\tilde{Y}_{u,t} - a - b\tilde{X}_{u,t})^2 K\left(\frac{u - t/T}{h}\right),$$

with respect to a, b. The first-order conditions are

$$\sum_{t=1}^{T} K_{ut}\tilde{Z}_{u,t}(\tilde{Y}_{u,t} - \tilde{Z}_{u,t}^\top \tilde{\vartheta}(u)) = 0. \tag{12}$$

$$\tilde{\vartheta}(u) = \left[\sum_{t=1}^{T} K_{ut}\tilde{Z}_{u,t}\tilde{Z}_{u,t}^\top\right]^{-1} \sum_{t=1}^{T} K_{ut}\tilde{Z}_{u,t}\tilde{Y}_{u,t},$$

[1] This method is a modification of what Aït-Sahalia (1996) proposed in an attempt to address time-varying nature of the diffusion process of interest. As suggested in Aït-Sahalia (1996), this method doesn't guarantee full efficiency. However, Aït-Sahalia (1996) suggested a method to improve efficiency. That method can also be applied here. In addition, there is an alternative way. Kristensen (2004) proposed a procedure of estimation of more general form of the drift term. We believe his method could be used as well. However, since we restrict our attention to the specification (8), his method is not discussed here.

where $\tilde{\vartheta}(u) = [\tilde{a}, \tilde{b}]^\top$, $\tilde{Z}_{u,t}$ =$[1, \tilde{X}_{u,t}]$. Meanwhile, in the neighborhood of a fixed time point u, the Kolmogorov backward equation yields [2]

$$E\left[\tilde{X}_{u,t+\Delta}|\tilde{X}_{u,t}\right] = \alpha + e^{-\Delta\beta}\left(\tilde{X}_{u,t} - \alpha\right). \tag{13}$$

Comparing the above equation (13) with the previous equation (11) shows the following relationships; $\tilde{\theta} = [\tilde{\alpha}, \tilde{\beta}]^\top$ where $\tilde{\alpha} = -\frac{\tilde{a}}{\tilde{b}}$ and $\tilde{\beta} = -\ln(\tilde{b}+1)/\Delta$. Likewise, the corresponding estimator from available data generated from the process $\{X_t\}$ can be

$$\hat{\vartheta}(u) = \left[\sum_{t=1}^{T} K_{ut} Z_t Z_t^\top\right]^{-1} \sum_{t=1}^{T} K_{ut} Z_t Y_t, \qquad \text{where } \hat{\vartheta} = [\hat{a}, \hat{b}]^\top. \tag{14}$$

where $Y_t = X_{t+\Delta} - X_t$. Let $\vartheta = [a, b]^\top$ and $Z_t = [1, X_t]^\top$. We propose our estimators of θ as $\hat{\alpha} = -\frac{\hat{a}}{\hat{b}}$ and $\hat{\beta} = -\ln(\hat{b}+1)/\Delta$. Therefore, from $\hat{\vartheta}$, we can indirectly acquire $\hat{\theta} = [\hat{\alpha}, \hat{\beta}]^\top$. Due to Proposition 1, $\tilde{\vartheta}(u)$ and $\hat{\vartheta}(u)$ share the same asymptotic properties and so do $\hat{\theta}$ and $\tilde{\theta}$. Consequently, the estimator of the drift term can be expressed as

$$\hat{\mu}(u, X_{u,t}; \hat{\theta}) = \widehat{\beta}(u)(\widehat{\alpha}(u) - X_{u,t}).$$

4.2.2 The Diffusion

The following argument is based on our discussion in section (4.1.2), and analogous to that in Aït-Sahalia (1996). To begin with, let $f_u(\cdot)$ and $p_u(\cdot)$ denote the marginal density of the spot rate and the transitional density between two contiguous $\tilde{X}_{u,t}$, $\tilde{X}_{u,t+\Delta}$ respectively. Under the assumption of local stationarity, the Kolmogorov forward equation can be modified as

$$\begin{aligned}\frac{\partial p_u(\tilde{X}_{u,t+\Delta}, \Delta|\tilde{X}_{u,t}, u)}{\partial \Delta} &= -\frac{\partial}{\partial \tilde{X}_{u,t+\Delta}}(\mu_u(\tilde{X}_{u,t+\Delta}; \theta) p_u(\tilde{X}_{u,t+\Delta}, \Delta|\tilde{X}_{u,t}, u)) \\ &+ \frac{1}{2}\frac{\partial^2}{\partial \tilde{X}^2_{u,t+\Delta}}(\sigma_u^2(\tilde{X}_{u,t+\Delta}) p_u(\tilde{X}_{u,t+\Delta}, \Delta|\tilde{X}_{u,t}, u)).\end{aligned}$$

[2] Let $\varphi_u(\tilde{X}, t)$ be the solution of the backward Kolmogorov equation $\partial\varphi_u(\tilde{X}, t)/\partial t = A\varphi_u(\tilde{X}, t)$ with initial condition $\varphi(\tilde{X}, 0) = \tilde{X}_0 = X_0$, where A is the backward Kolmogorov operator: $A\varphi_u(\tilde{X}, t) \equiv \mu_u(\tilde{X}, t, \theta)/\partial\varphi_u(\tilde{X}, t)\partial X + (1/2)\sigma_u^2(\tilde{X}, t)\partial^2\varphi_u(\tilde{X}, t)/\partial\tilde{X}^2$. This partial differential equation has the unique solution $\varphi_u(\tilde{X}, t) = E_u[\tilde{X}_t|\tilde{X}_0]$, by Dynkin's formula. Also, the function $\psi_u(\tilde{X}, t) \equiv \alpha + e^{-\beta\Delta}\left(\tilde{X}_t - \alpha\right)$ also satisfies the equation with the same initial equation. Thus $\psi = \varphi$. Since we assume local stationarity, we get the result (13).

Multiplying $f_u(\tilde{X}_{u,t},\Delta)$ on both sides and integrating both sides over $[0,\infty)$,

$$\begin{aligned}&\int_0^\infty \frac{\partial p_u(\tilde{X}_{u,t+\Delta},\Delta|\tilde{X}_{u,t},u)}{\partial\Delta} f_u(\tilde{X}_{u,t},\Delta)d\tilde{X}_{u,t} = \\ &\int_0^\infty -\frac{\partial}{\partial\tilde{X}_{u,t+\Delta}}(\mu_u(\tilde{X}_{u,t+\Delta};\theta)p_u(\tilde{X}_{u,t+\Delta},\Delta|\tilde{X}_{u,t},u))f_u(\tilde{X}_{u,t},\Delta)d\tilde{X}_{u,t} \quad + \\ &\frac{1}{2}\int_0^\infty \frac{\partial^2}{\partial\tilde{X}^2_{u,t+\Delta}}(\sigma_u^2(\tilde{X}_{u,t+\Delta})p_u(\tilde{X}_{u,t+\Delta},\Delta|\tilde{X}_{u,t},u))f_u(\tilde{X}_{u,t},\Delta)d\tilde{X}_{u,t}.\end{aligned} \tag{15}$$

Since a partial derivative of the equation (16) with respect to time is zero in the vicinity of u due to local stationarity, the left hand side of (15) is zero, and

$$\int_0^\infty p_u(\tilde{X}_{u,t+\Delta},u+\Delta|\tilde{X}_{u,t},u)f_u(\tilde{X}_{u,t})d\tilde{X}_{u,t} = f_u(\tilde{X}_{u,t+\Delta},\Delta). \tag{16}$$

As a result, we obtain

$$\frac{1}{2}\frac{\partial^2}{\partial\tilde{X}^2}(\sigma_u^2(\tilde{X})f_u(\tilde{X})) = \frac{\partial}{\partial\tilde{X}}(\mu_u(\tilde{X};\theta)f_u(\tilde{X})). \tag{17}$$

The above equation is valid over a range in which the process behaves in a stationary manner, say in the neighborhood of a given time point u. Finally, from (17) and Proposition 1, the diffusion term can be obtained as

$$\hat{\sigma}_u^2(x;\hat{\mu},\hat{f}_u(x)) = \frac{2}{\hat{f}_u(x)}\int_{-\infty}^x \hat{\mu}_u(y;\hat{\theta})\hat{f}_u(y)dy.$$

5 Asymptotics

This section establishes asymptotic properties of our estimators proposed in the previous section. It is important to bear in mind that Proposition 1 and Proposition 2 allows us to analyze the asymptotic properties of our estimators in the vicinity of the time point u. We show that our estimators are pointwise consistent and asymptotically normal. To this end, we show our estimators are weak and mean square error consistent. Furthermore, we provide the uniform convergence rate of our estimators of the drift and the diffusion function. In the following, μ_0, σ_0^2, g_0, and f_0 denote the true drift, diffusion, the density function of $\{X_t\}$ and the density function of $\{\tilde{X}_{u,t}\}$.

Given that the diffusion function is obtained from density matching method, our asymptotic analysis starts from that of kernel estimator of the pdf in the vicinity of a certain time point u. Since we obtain our estimator of the density nonparametrically, we set out regularity conditions for a kernel function and bandwidths as well as those for the existence of the density function $f_0(u,x)$ before we proceed with asymptotic theory of our estimators.

Assumption 2 *(i) For* $\forall x\in\mathbf{I}$ *and* $\forall u\in[0,1]$, $\mu_0(u,x)$, $\sigma_0^2(u,x)\in\mathbb{C}_2(C_0)$ *with* $\sigma_0^2(u,x)>0$; *(ii)* $\inf_{u\in[0,1]}\beta(u)>0$.

Assumption 2 is required to ensure the existence of a unique solution to (3), the validity of the relationship among the density, the drift and the diffusion, and the existence of a strictly stationary density function associated with $\{\tilde{X}_{u,t}\}$. In particular, Assumption 2 implies C1 - C5 under our parameterization of the drift function as in (8) (see Kristensen (2004)). For our asymptotic analysis, we only require that $\{\tilde{X}_{u,t}\}$ are α-mixing. Note that β-mixing is stronger than α-mixing.

Assumption 3 (Kernel) *The kernel $K(\cdot)$ is a bounded symmetric around zero function such that: (i) it is continuously differentiable up to order r on $\mathbb{R}$ with $2 \leq r$; (ii) it belongs to L^2, $\int |K(x)|dx < \infty$, $\int K(x)dx = 1$, and the support of K is contained in $[-1, 1]$; (iii) $\mu_i(K) = \int x^i K(x)dx = 0$, $i = 1, \ldots, r-1$, and: $\int x^r K(x)dx \neq 0$, $\int |x|^r |K(x)|dx < \infty$, $\lim_{||u|| \to \infty} ||u|| K(u) = 0$; (iv) $K(\cdot)$ is Lipschitz continuous, i.e. $|K(u) - K(u')| \leq C_3 |u - u'|$ for all $u, u' \in \mathbb{R}^2$.*

The bigger is r, the more smooth is the density f. Also, a higher order kernel is ideal when it comes to obtaining the better rates of convergence. For our asymptotic analysis, we assume $r = 2$. Only K for the time index u needs to be confined to $[-1, 1]$.

The above Assumption 3 is standard in much of the nonparametric literature.[3]

Assumption 4 (Bandwidth Choices) *Let h be the bandwidth for the drift estimation and h_1 and h_2 be the bandwidths for the density and volatility estimation. Also, note that h_1 is used for the time point argument u whereas h_2 is used for the state variable argument x. (i) to estimate μ: as $T \to \infty$, $h \to 0$, and $Th \to \infty$; (ii) to estimate f, σ^2: as $T \to \infty$, $\max(h_1, h_2) \to 0$ and $Th_1h_2 \to \infty$.*

Theoretical optimal bandwidths should reflect the local segment length of time on which assumption of stationarity is valid. In particular, the bandwidth, $h_{1,opt}$ is important since it is closely related with the vicinity of a time point where stationarity holds true. This is because the interval of time homogeneity of unknown size possibly varies over time and therefore different bandwidths could be used for different time points. In particular, it can be shown that $h_{opt} = \Lambda(t/T, \theta) T^{-1/5}$ in MSE/IMSE sense. Similar argument applies to $h_{1,opt}$ and $h_{2,opt}$. See Müller and Stadtmüller (1987) and Cai (2007). Therefore, it is natural that we should consider time varying bandwidths since $\Lambda(\cdot)$ depends on time let alone the unknown parameters. However, we don't delve into the optimal bandwidth selection theoretically in this paper.

5.1 Consistency

In this section, we show weak consistency results of our estimators.

[3] Hansen (2008) and Kristensen (2009) showed that the following can be assumed for uniform convergence instead of Lipschitz continuity of K.

$$|K^{(1)}(u)| \leq C||u||^{-v} \text{ for some } v > 1.$$

Theorem 1 *Suppose that Assumptions 1 - 4 hold. Let u and x be interior points over their domain. Then, $\hat{\mu}$, $\hat{\sigma}^2$, the nonparametric estimators of the drift and the diffusion functions of the underlying process as well as the density estimator $\hat{f}$ are weakly consistent. Specifically, as $T \to \infty$:*

$$\hat{f}\,(u,x) \xrightarrow{p} g_0\,(u,x) \tag{18}$$

$$\hat{\mu}(u,x;\hat{\theta}) \xrightarrow{p} \mu_0\,(u,x;\theta_0) \tag{19}$$

$$\hat{\sigma}^2(u,x;\hat{\mu},\hat{f}) \xrightarrow{p} \sigma_0^2\,(u,x;\mu_0,f_0)\,. \tag{20}$$

As is the case with density estimators with dependent data, our approach to MSE of $\hat{f}(u,x)$ involves a breakdown into short range and long range to bound covariances. On the other hand, recall that the drift, μ is parameterized up to an unknown parameter function vector θ, which can be derived from ϑ by the Kolmogorov backward equation. Given ϑ is estimated by local regression method, we can show the consistency result of θ by showing that of ϑ due to the Slutsky theorem. In addition, the diffusion estimator, $\hat{\sigma}^2$ has the same rate of pointwise convergence as the marginal density estimator $\hat{f}$.

5.2 Asymptotic Normality

In this section, we present the asymptotic normality of the estimators for the drift and diffusion functions. To begin with, we assume the following assumptions hold for asymptotic normality results.

Assumption 5 *The density f_{u,t_1,t_2,t_3,t_4} of $\left(\tilde{X}_{u,t_1},\ldots,\tilde{X}_{u,t_4}\right)$ exists whenever $t_1 < t_2 < t_3 < t_4$ and $\sup_{t_1<t_2<t_3<t_4} ||f_{u,t_1,t_2,t_3,t_4}||_\infty < \infty$.*

Assumption 6 $\sup_{t\geq 1} E[X_t]^{4+\eta} < \infty$ *for some positive η.*

To present the asymptotic normality, we require some additional notation. Let $x_1,\ldots,x_k$ be distinct points, and let

$$\text{bias}_{\hat{\theta}}(u) = \mu_2(K)\Gamma_\theta(u)M_u^{-1}E\left[\tilde{Z}_{u,t}\tilde{Z}_{u,t}^\top\left\{\vartheta_0^{(1)}\,(u)\left[\frac{\dot{f}_0\left(u,\tilde{X}_{u,t}\right)}{f_0\left(u,\tilde{X}_{u,t}\right)}\right]+\frac{1}{2}\vartheta_0^{(2)}\,(u)\right\}\right],$$

where $\tilde{Z}_{u,t} = [1,\tilde{X}_{u,t}]$, $M_u := E[\tilde{Z}_{u,t}\tilde{Z}_{u,t}^\top]$, and $\Omega_u := E[\tilde{Z}_{u,t}\tilde{Z}_{u,t}^\top\sigma_u^2(\tilde{X}_{u,t})]$, while:

$$\text{bias}^1_{\hat{\sigma}^2}(u,x_i) = \frac{\mu_2(K)}{2}\sigma_0^2\,(u,x_i)\,\frac{\ddot{f}_0(u,x_i)}{f_0(u,x_i)}$$

$$\text{bias}^2_{\hat{\sigma}^2}(u,x_i) = \frac{\mu_2(K)}{2}\sigma_0^2\,(u,x_i)\,\frac{f_0^{(2)}(u,x_i)}{f_0(u,x_i)}$$

$$V_\theta(u) = ||K||_2^2\Gamma_\theta(u)M_u^{-1}\Omega_u M_u^{-1}\Gamma_\theta^\top(u)$$

$$V_\sigma = \operatorname{diag}\left\{V_\sigma(u, x_i), i = 1, \ldots, k\right\}, \quad V_\sigma(u, x_i) = ||K||_2^4 \frac{\sigma_0^4(u, x_i)}{f_0(u, x_i)}.$$

$$\Gamma_\theta(u) = \frac{\partial \theta}{\partial \vartheta} = \begin{pmatrix} -\dfrac{1}{b(u)} & \dfrac{a(u)}{b(u)^2} \\ 0 & -\dfrac{1}{b(u)+1} \end{pmatrix},$$

where $\theta = [\alpha, \beta]^\top$ and $\vartheta = [a, b]^\top$.

Theorem 2 *Suppose that the Assumptions 1 - 6 hold. Let u and x be interior points over their domain. Then $\hat{\theta}$ and $\hat{\sigma}^2$ are pointwise asymptotically normally distributed. In addition, for distinct time points u, x and u^*, x^*, we have $\hat{\theta}(u)$, $\hat{\sigma}^2(u, x)$ are asymptotically independent of $\hat{\theta}(u^*)$, $\hat{\sigma}^2(u^*, x^*)$ respectively. Specifically, we have:*

$$\sqrt{Th}\left\{\hat{\theta}\left(u\right) - \theta_0\left(u\right) - h^2 \text{bias}_{\hat{\theta}}(u)\right\} \quad \xrightarrow{d} \quad N\left(0, V_\theta(u)\right), \tag{21}$$

$$\sqrt{Th_1 h_2}\left\{\hat{\sigma}^2(u, x_i) - \sigma_0^2(u, x_i) - h_1^2 \text{bias}^1_{\hat{\sigma}^2}(u, x_i) - h_2^2 \text{bias}^2_{\hat{\sigma}^2}(u, x_i)\right\}_{i=1}^k \quad \xrightarrow{d} \quad N(0, V_\sigma). \tag{22}$$

Consistent estimators of the asymptotic variance for $\hat{\theta}\left(u\right)$ and $\hat{\sigma}^2(u, x)$ can be constructed respectively as

$$\begin{aligned} \hat{V}_\theta\left(u\right) &= ||K||_2^4 \hat{\Gamma}_\theta(u) \hat{M}_u^{-1} \hat{\Omega}_u \hat{M}_u^{-1} \hat{\Gamma}_\theta^\top(u) \\ \hat{V}_\sigma\left(u, x_i\right) &= ||K||_2^4 \hat{\sigma}^4\left(u, x_i\right) / \hat{f}\left(u, x_i\right) \text{ for } i = 1, \ldots, k, \end{aligned}$$

where $\hat{M}_u = (Th)^{-1} \sum_{t=1}^T Z_t Z_t^\top K_{ut}$, $\hat{\Omega}_u = (Th)^{-1} \sum_{t=1}^T Z_t Z_t^\top \hat{v}_t^2 K_{ut}^2$ and $\hat{v}_t = y_t - Z_t^\top \hat{\vartheta}\left(u\right)$, while

$$\hat{\Gamma}_\theta(u) = \begin{pmatrix} -\dfrac{1}{\hat{b}(u)} & \dfrac{\hat{a}(u)}{\hat{b}^2(u)} \\ 0 & -\dfrac{1}{\hat{b}(u)+1} \end{pmatrix}.$$

Note that the overall optimal bandwidths are $h = O(T^{-1/5})$, while $h_1, h_2 = O(T^{-1/6})$ as mentioned before in conjunction with Assumption 4 even though time varying bandwidths are much more appropriate. This results in asymptotic mean squared errors of order $T^{-4/5}$ and $T^{-2/3}$ respectively.

5.3 Uniform Convergence Rates

This section proposes uniform convergence of our nonparametric estimators of the drift and the diffusion functions. Whereas most of papers whose focus was non- or semiparametric estimation of the diffusion processes showed the pointwise consistency and asymptotic normality of their proposed estimators, a few sought after the uniform consistency results. Our approach to uniform convergence is based on the methodologies of Hansen (2008) and Kristensen (2009). In previous sections, we restrict our attention to the vicinity of a fixed time point, u. However, in this section, we deal with heterogeneous dependent data over the whole span. In particular, we require the data to be strong mixing even if not identically distributed. The main idea is instead of assuming stationarity and hence dividing the compact domain into with the same length, we bound the biggest sized cube of appropriately divided cube set of compact domain. See Kristensen (2009).

Assumption 7 *The triangular array $\{X_t\}_{t=1}^{T}$ are strongly mixing with its mixing coefficients $\alpha_{j,T}$ such that for $\rho > 0$,*

$$\alpha_{j,T} = \sup_{=0<j\leq T} \sup_{A\in\mathcal{F}_{-\infty}^{t}, B\in\mathcal{F}_{t+j}^{\infty}} \left|P\left(A\cap B\right) - P\left(A\right)P\left(B\right)\right| \leq Cj^{-\rho}$$

where $\alpha_{j,T} \to 0$ as T, $j \to \infty$.

Assumption 8 *The density of $\{X_t\}$, $g_t\left(x\right)$ and joint densities of $\left(X_t, X_{t+j}\right)$, $g_{t,t+j}\left(x,y\right)$ are uniformly bounded.*

Theorem 3 *Suppose Assumptions 1 - 8 hold. Furthermore, suppose that*

$$r\left(T\right) := \left(\frac{\ln T}{Th_1h_2}\right)^{1/2} = o\left(1\right), \; r_*\left(T\right) := \left(\frac{\ln T}{Th}\right)^{1/2}$$

Then, for any sequence δ_T such that $\delta_T/h_1 \to \infty$ and $\delta_T \to 0$, we have

$$\begin{aligned}
\sup_{x\in I, u\in[\delta_T, 1-\delta_T]} \left|\hat{f}\left(u,x\right) - g_0\left(u,x\right)\right| &= O_p\left(h_1^2 + h_2^2\right) + O_p\left(r\left(T\right)\right) \\
\sup_{u\in[\delta_T, 1-\delta_T]} \left|\hat{\alpha}\left(u\right) - \alpha_0\left(u\right)\right| &= O_p\left(h^2\right) + O_p\left(r_*\left(T\right)\right) \\
\sup_{u\in[\delta_T, 1-\delta_T]} \left|\hat{\beta}\left(u\right) - \beta_0\left(u\right)\right| &= O_p\left(h^2\right) + O_p\left(r_*\left(T\right)\right) \\
\sup_{x\in I, u\in[\delta_T, 1-\delta_T]} \left|\hat{\sigma}^2\left(u,x\right) - \sigma_0^2\left(u,x\right)\right| &= O_p\left(h_1^2 + h_2^2\right) + O_p\left(r\left(T\right)\right).
\end{aligned}$$

As we noted before, we estimated the drift function via a local constant least squares method. However, a local constant least squares method is known to be subject to boundary bias problems and therefore, we avoid this difficulty by considering uniform convergence over the interval $u \in [\delta_T, 1-\delta_T]$ where δ_T is any sequence such that $\delta_T/h_1 \to \infty$ and $\delta_T \to 0$. In general, it is known that the boundary bias problems can be avoided by adopting a local linear or polynomial least squares method. Although we could use those methods here, it is not an issue of this paper.

Note that the uniform convergence rate of the diffusion function is dominated by that of the density function.

6 Application

In this section, we apply our estimation procedure to the real data in order to illustrate the validity of our methodology. Moreover, we calculate prices of a zero coupon bond and its call option by using our estimates of functions of interest. Under the usual mathematical finance, no-arbitrage condition and hedging determines the price of derivatives associated with underlying assets. Given the time varying diffusion processes specified as in (2), we provide estimates of option prices using our estimates of the drift and diffusion functions. We restrict our focus on interest-rate-derivative securities. The time series plots of the underlying interest rates for option pricing are shown in the following Figure 2.

FIGURE 2 ABOUT HERE

Throughout our application, the Epanechnikov kernel $K(v) = 0.75(1 - v^2)I(|v| \leq 1)$ is used for the time point argument, u and the Gaussian kernel for the state variable argument, x. For the bandwidths, we used the Silverman's rule of thumb i.e. $h = \hat{SD}(u)T^{-1/5}$ and h_1, $h_2 = \hat{SD}(u)T^{-1/6}$, $\hat{SD}(X_t)T^{-1/6}$ respectively. Application results are shown in Appendix.

In the application, we compare our results with those of Aït-Sahalia (1996) on the same data.[4] In Aït-Sahalia (1996), due to the assumption of stationarity, while parameters $\theta = [\alpha, \beta]^\top$ are fixed, the volatility function only depends on the state variable x. By contrast, in our model, θ are time varying and the volatility depends on time as well as the state variable, the interest rate, x.

6.1 Data

Even though Aït-Sahalia (1996) proposed nice methodology to estimate the drift and diffusion functions from discretely sampled data, there has been one important criticism regarding his method. That is, the actual short run interest rates might not be stationary. It may well be nonstationary which renders his method invalid. For example, it might contain a unit root and hence the method based on stationarity assumption over the whole period might not go through. Quite a few papers studied whether asset prices including interest rates follow unit root processes. Although a clear cut answer on that issue has not been provided, it would be ideal if our method based on the local argument could alleviate the degree of unit rootness embedded in the data to which the methodology is applied. In this section, to shed some light on this issue, we provide the results of several unit root tests on the data we used for our application. There are an array of unit root tests. However, several celebrated unit root tests such as standard Dickey Fuller (ADF) and Phillips-Perron (PP) don't necessarily suit the short run interest rates data whose DGP is assumed to be continuous-time diffusion process. This is mainly because the conditional variance, possibly as a function of time and its own state variable, tends to vary across time and display heteroscedasticity and therefore, the constant variance assumption of those unit root tests is obviously unsatisfied. A collection of econometric literature has documented considerably low power of standard unit root tests against stationary alternative when some of their assumptions are violated. Under such circumstances,

[4] For the data description, see Aït-Sahalia (1996). Kristensen and Aït-Sahalia gave advice on the data acquisition. Especially, Kristensen gave advice on option pricing as well. We take this opportunity to thank them for their support.

those tests may well be unreliable, even though they are widely used in most of empirical literature on asset prices. In the absence of pre-knowledge of the drift and diffusion functions of the process, there is high risk of misspecification of variance and those widely used unit root tests are subject to low power and might lead to a false conclusion. Quite recently though, a series of attempts have been made to ensure robustness of tests against non-constant variance. Breitung (2002) proposed a unit root test which does not require the specification of the short-run dynamics or the estimation of nuisance parameters along with tabulated critical values with selected significant levels. Main idea of his unit root test is the asymptotic theory for a unit root test based on ranks does not involve the parameters involved by the short run dynamics of the process. It was shown that the nonparametric test is robust to GARCH errors even under integrated or explosive volatility, which could be made to converge to the diffusion process. Cavaliere (2004) also delved into the unit root tests under time varying variances. It is worth noting that several papers have considered the performance of several unit root tests under various settings and suggested Breitung's nonparametric unit root test is more robust to misspecification and time varying variance than any other unit root tests. In this regard, we use Breitung's nonparametric unit test (NP) in order to confirm whether the data has a unit root. We also provide PP for the comparison purpose. However, appropriate caution should be exerted since to the best of our knowledge, there is no unit root test specifically designed for the data generated from continuous time diffusion processes even though unit root tests for GARCH errors have been delved into.

To begin with, we use PP test statistic as

$$PP = (s_u/s_T)\, t_\beta - (1/2)\left(s_T^2 - s_u^2\right)\left\{s_T\left[T^{-2}\sum_{t=1}^{T}\left(X_{t-1}-\bar{X}\right)^2\right]^{1/2}\right\}^{-1},$$

where $T^{-1}\sum \hat{u}_t^2 + 2T^{-1}\sum_{j=1}^{T-1} K\left(\frac{j}{h}\right)\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}$ and $\hat{u}_i = y_i - \bar{y}$ and $\bar{y}$ is the sample mean. Meanwhile, NP test statistic is

$$NP = \frac{T^{-2}\sum_{t=1}^{T}\hat{U}_t^2}{\sum_{t=1}^{T}\hat{u}_t^2}$$

where $\hat{U}_t = \sum_{i=1}^{t}\hat{u}_i$, $\hat{u}_i$ is defined above. The null of a unit root is rejected if the value of test statistic is smaller than the corresponding critical value.

Table 1 shows results of two types of unit root tests described above over the different sample period. We chose five different sub sample periods. Critical values associated with several selected significance levels are also provided. Table 1 shows local argument alleviates the unit rootness by lowering values of test statistics for PP and NP.

TABLE 1 ABOUT HERE

6.2 Results of estimation of the drift and diffusion

Recall that in our specification of the drift term, α can be considered as the equilibrium level or a steady state mean of the interest rate concerned and β represents the rate of mean reverting adjustment. In Figure 3.(a) and 3.(b), our estimates of $\alpha(u)$ and $\beta(u)$ and those of α (=0.083082) and β (=1.6088) in Aït-Sahalia (1996) are plotted together respectively for the purpose of comparison. It is worth noting that our estimation methodology is well suited for both stationary and nonstationary, in particular time varying processes. More specifically, the estimators of the drift and the diffusion functions proposed by Aït-Sahalia (1996) can be easily obtained by setting the bandwidth in relation to the time argument u infinity in our methodology. Figures 3.(c) shows our density estimates under the local stationarity. Meanwhile, the first two rows of Figure 4 explore the change of density function and our diffusion function over time. We can compare our estimates of density and diffusion functions at different times with those of Aït-Sahalia (1996) shown in the first column of Figure 4. The second, third and the fourth columns of Figure 4 depict the estimated curves of our density function $\hat{f}$ and $\hat{\sigma}^2$ at different times. Also, along with those estimated curves, we plot their corresponding 99% pointwise confidence intervals from the limit distribution given in Section 5 without the bias correction.

Given that our time varying α and β are often out of the 99% confidence interval of the fixed α and β in Aït-Sahalia (1996), stationarity assumption with respect to the fixed α and β seems too strong. Jagged shape shown in our estimate of β may indicate possible need of variable bandwidths. The shape of the density function varies slightly at each different time but doesn't change drastically whereas the shape of the diffusion function changes over time.[5] In particular, changes of the shape of the diffusion function over time are noticeable. It is consistent with empirical literature which argues that volatility changes over time. On balance, the drift and diffusion functions are very much time dependent and our methodology captures those properties so that it can complement Aït-Sahalia (1996) quite well.

In sum, results of our real data application is consistent with the asymptotic results given in Section 5. There are several things worth noting. As is known with other nonparametric estimation methods, the choice of our kernel functions is not so relevant. Meanwhile, much more attention should be paid to the choice of bandwidths, especially the one for a time argument u. For our asymptotic analysis, we only require Assumption 4. However, for the finite sample applications, it would be better for the bandwidth of a time argument u to reflect the local time interval in which the process is stationary in order to obtain better results and optimal convergence rates as we mentioned before. Moreover, it would be interesting to find an automatic data-driven algorithm for the variable bandwidths which should be compatible with Assumption 1. This doesn't seem to be obvious but challenging. Therefore, we leave this topic to the future research.

[5] The change of shape of density functions may be partly due to the difference of number of observations available over a certain area of state value, X. Since we don't have many data available for the high interest rate area, the vertical section sliced at the time point in the high interest rate period might not show bell-shape density function whereas in any other areas where enough data reside, bell shape density function emerges even though it is a bit different from the normal density. Also, the change of those of volatility functions is also affected by the number of observations along with the change of the drift functions. We produced different pictures by slicing the three-dimensional density and volatility estimates at 20 different time points. Overall, those pictures show qualitatively similar patterns as in Figure 2.

6.3 Option pricing

As seen from a number of papers, (For extensive review, see Kristensen (2008)) our estimation of the drift and diffusion functions in the locally stationary models can be applied to asset pricing theory, in particular to the pricing of financial derivatives. In this section, we use our estimates of the drift and diffusion functions in the previous section in order to obtain implied prices of zero coupon bond (ZCB) and its call option (ZBC).[6] More specifically, we use the Feynman-Kac Representation provided below. It is worth mentioning that the process (2) is measurable with respect to the physical measure or real world measure P and therefore our estimators are measurable with respect to P whereas Feynman-Kac Representation involves the risk free measure or martingale equivalent measure Q.

Let $\mathcal{A}$ be the derivative price at time t with maturity T and underlying variable X_t such as the spot interest rate in our application.

$$\begin{aligned}\mathcal{A}(t,x) = E^{Q}\Bigg[& b\left(X_T\right)\exp\left\{-\int_t^T X_s ds\right\} \\ &+ \int_t^T \exp\left\{-\int_t^T X_s ds\right\} c\left(s, X_s\right) ds \Bigg| X_t = x\Bigg], \end{aligned} \tag{23}$$

where $b\left(X_T\right)$ is the payoff of the derivative at maturity, and $c\left(s, X_s\right)$ is the cash flow at time s.

Our estimation procedure for derivative prices is the Monte-Carlo simulations method as follows. First, from historical interest rate data, we obtain estimates $\left(\hat{\mu}, \hat{\sigma}^2\right)$. Then, we are in a position to simulate the sample paths of the following risk neutral process,

$$dX_t = \left[\mu\left(t/T, X_t\right) - \gamma\left(t/T, X_t\right)\sigma\left(t/T, X_t\right)\right]dt + \sigma\left(t/T, X_t\right)dW_t, \tag{24}$$

where $\gamma\left(t\right)$ is the time varying market risk premium of interest rates. We plug the obtained estimates into the drift and diffusion in (24) along with the estimate of γ in order to obtain the sample paths of $\{X_t\}$ under the risk neutral measure. Using (23) and $\{X_t\}$ under Q, the price $\mathcal{A}(t,x)$ can be estimated as a sample conditional mean over the simulated sample paths.

6.3.1 Market Price of Risk

It is important to estimate the market price of risk since we use the Feynman-Kac representation. The market price of risk, $\gamma\left(\cdot\right)$ is estimated nonparametrically along the lines of Stanton (1997). More specifically, since the market price of risk, $\gamma\left(\cdot\right)$ is thought of to be the extra compensation per unit of risk for taking on financial derivatives, the following approximation holds.

$$\gamma_u\left(X_t\right) \approx \frac{1}{\sigma_u^{[1]}\left(X_t\right) - \sigma_u^{[2]}\left(X_t\right)}\left[E_t\left(X_{t,t+1}^{[1]} - X_{t,t+1}^{[2]}\right)\Big| X_t\right], \tag{25}$$

[6] The rationale for our option pricing application is provided briefly in the Appendix C. For extensive details, see Hull and White (1990), Stanton (1997) and Kristensen (2008).

where $X_{t,t+1}^{[i]}$ denotes the interest rate of i asset for one unit of holding period. See Appendix C for more details. (25) can be estimated as

$$\hat{\gamma}(u,x) = \frac{1}{\hat{\sigma}^{[1]}(u,x) - \hat{\sigma}^{[2]}(u,x)} \frac{\sum_{t=1}^{T-1} \left(X_{t,t+1}^{[1]} - X_{t,t+1}^{[2]} \right) K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{x-X_t}{h_2}\right)}{\sum_{t=1}^{T-1} K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{x-X_t}{h_2}\right)}.$$

Results are provided in the third row of Figure 4.

6.3.2 Zero coupon bond pricing (ZCB)

To begin with, we start with option pricing of zero coupon bond. We consider the price of a discount bond whose underlying asset's current value and the bond's payoff is x and \$1 at expiry, T respectively. Consider the following function $\mathcal{A}_u(t,x)$. Also, Initial and boundary conditions can be given as follows.

$$\mathcal{A}_u(t,x) = E_{t,x}^{Q}\left[\exp\left(-\int_t^T X_s ds\right)\right], \tag{26}$$

where $E_{t,x}^{Q}[\cdot] = E[\cdot|X_t = x]$ under the risk neutral measure Q. Note that $c(s, X_s) = 0$ for $\forall s \in [0,T]$ and $b(X_T) = 1$. Results are provided in Table 2.

6.3.3 European call option (ZBC)

As far as a European call option whose underlying asset is a zero coupon bond is concerned, we consider the following.

$$\mathcal{A}_u(t,x) = E_{t,x}^{Q}\left[\max\left(0, \mathbb{B}(x, S-T, S) - \mathcal{X}\right) \exp\left(-\int_t^T X_s ds\right)\right],$$

where $\mathbb{B}$ is the price of underlying zero coupon bond at the maturity of the call option of interest and $\mathcal{X}$ is the strike price. Results are provided in Table 3.

7 Simulation study

In this section, we provide simulation results to examine the finite-sample performance of our estimators. Simulation results are provided in Appendix. We consider four different models to investigate the robustness of our estimators.

7.1 Simulated Models

In order to conduct a Monte-carlo experiment, we focus on the simplified version of Hull and White model (HW, 1990), the extended version of the Cox, Ingersoll, and Ross model (CIR model).

$$dX_t = \beta(t)(\alpha(t) - X_t)dt + \sigma(t)X_t^i dW_t, \quad i = 1/2.$$

Whereas the CIR model doesn't capture the time varying features of the drift and the diffusion we have a keen interest in, the HW model incorporates time-dependent properties of them. Following Euler discretization scheme, observations are created from each discretized version which corresponds to its continuous one by leaving 100 realizations unobserved between observed ones. The following specifications are what we simulated from. To begin with,

Example 1

$$dX_t = \beta(u)(\alpha(u) - X_t)dt + \tau(0.5\cos(0.5\pi u) + 1)\sqrt{X_t}dW_t$$

Example 2

$$dX_t = \beta(u)(\alpha(u) - X_t)dt + \tau 0.5(\cos(0.5\pi u) + (\sin(0.5\pi u)))\sqrt{X_t}dW_t$$

Example 3

$$dX_t = \beta(\alpha - X_t)dt + \tau(0.25u + 1)\sqrt{X_t}dW_t,$$

where $u = t/T$, $\alpha(u) = -u^2 + u$, $\beta(u) = 0.2146$, $\tau = 0.0783$ *in all of the above examples.*

Secondly, we include a time-homogeneous case (CIR model) to investigate whether our methods contain an important trait of verifying the correct structure of the model. That is, it would be good if the proposed methods in this paper could identify both quantities of interest, whether the process is stationary or locally stationary. The fourth example will serve this purpose.

Example 4

$$dX_t = \beta(\alpha - X_t)dt + \tau\sqrt{X_t}dW_t,$$

where $\alpha = 0.0857$, $\beta = 0.2146$, *and* $\tau = 0.0783$.

These values are taken from Chapman and Pearson (2000). For simplicity, $\Delta = 1$. We conducted 500 simulations of each example with sample size $T = 1000, 3000$. If we presume we deal with weekly data, for example, T=1000 corresponds to around 20 years. For the kernel function with respect to u and x, we employ an Epanechnikov kernel where $K(u) = \frac{3}{4}(1 - u^2)I(|u| \leq 1)$ where I is a indicator function. For the bandwidths, we used the Silverman's rule of thumb i.e. $h = \hat{SD}(u)T^{-1/5}$ and h_1, $h_2 = \hat{SD}(u)T^{-1/6}$, $\hat{SD}(X_t)T^{-1/6}$ respectively. Our selections of a kernel function and a bandwidth satisfy the previous Assumptions 3 and 4 in Section 5 respectively. Also, given that our simulated model is quite analogous to CIR model, we chose initial values from an appropriate Gamma distribution.[7]

7.2 Simulation results

As is proven in Section 5, our simulation study lends evidence to our main results in Section 5. Our results confirm that the proposed estimators perform well with suitable sample size. More

[7] Initial values have the Gamma distribution $G(\nu, \lambda)$ where $\nu = \frac{2\alpha\beta}{\tau^2}$ and $\lambda = \frac{\tau^2}{2\beta}$.

specifically, Tables 4 and 5 in Appendix show that our estimators of the drift term and diffusion term perform well in pointwise MSE and integrated MSE senses respectively. Table 4 shows the pointwise performance of our estimators when $x\,(= .24 \text{ or } .08)$ and $u\,(= 1/2)$ are fixed at a certain value. Table 5 shows integrated MSE of our estimators. Even with a relatively reasonable sample size, the estimates of α and β from the local constant least regression method show their consistency. It is worth noting that there arises a boundary bias issue since we use the local constant regression instead of local linear one. This is a standard problem of kernel estimation methods and it is well-known that this boundary issue can be avoided by local linear implementation. Our diffusion estimator, $\sigma^2\,(u, x)$ also captures the true functional form of the simulated model quite well. The time nature in drift and volatility in each example is well captured by our estimators. Also, our method is robust to the change of the functional form of the diffusion and enables us to verify whether the structure of the underlying data generating process is stationary or locally stationary.

However, there are limitations we should be cautious about. With respect to the state variable domain, for relatively too small or too big values the state variable can take, our estimators of the volatility seem to perform less than the middle values which recur quite often. As we mentioned before, this is due to a standard problem of kernel estimation methods.

8 Concluding Remarks

In this paper, we propose an estimation method in an attempt to capture the time varying properties of the diffusion processes. We indirectly obtain the estimator of the drift term via local constant least squares, whereas we utilize the density matching method to obtain our estimator of the volatility of the process. It was shown that our estimators of time-inhomogeneous diffusion processes are consistent and asymptotically normally distributed under the assumption of local stationarity. Furthermore, those estimators uniformly converge to the true functions of the drift and the volatility. We provide applications with the real data as well as simulation results to illustrate our estimation procedure and validity of our theories. Our empirical results lend credence to our asymptotic theories. Our estimation procedure broadens the pool of estimation methodology of the diffusion processes by augmenting various estimation procedures under the assumption of stationarity.

There are many other challenges lying ahead, however. To begin with, pure diffusion processes have been under scrutiny since they don't seem to be compatible with several well-known stylized facts such as discontinuities of asset prices movement. It would be an important extension if we could allow for jump components under the similar setting. Possibility lies either in structural breaks in the functions of interest in time varying diffusion processes or in time varying jump-diffusion processes. Secondly, although the automated optimal variable bandwidth selection could be worth investigating, it is not discussed in this paper. Thirdly, since it is shown that our estimators of the drift and the diffusion function are asymptotically normally distributed, we could propose a test statistic for a nonparametric time-homogeneous diffusion model against our semiparametric time-inhomogeneous alternative among an array of possible tests. The corresponding theoretical justification could be worth delving into for a practical reason. Lastly, in this paper, we did not seek for efficiency improvement, even though it could be an interesting topic. Those topics are left as future research.

References

Aït-Sahalia, Y. 1996. Nonparametric Pricing of Interest Rate Derivative Securities. Econometrica, Vol.64, No. 3, 527-560.

Altman, N., 1990. Kernel Smoothing of Data With Correlated Errors. Journal of the American Statistical Association, Vol.85, No. 411, 749-759.

Andrews, D.W.K., 1995. Nonparametric Kernel Estimation for Semiparametric Models. Econometric Theory 11, 560-596.

Bandi, F.M., Phillips, P. C. B., 2003. Fully Nonparametric Estimation of Scalar Diffusion Models. Econometrica 71, 241-283.

Bandi, F. M., Nguyen, T., 2003. On the Functional Estimation of Jump-Diffusion Models. Journal of Econometrics, 2003, 116, pp. 293-328

Black, F., Scholes, M., 1973. The pricing of options and corporate liabilities. Journal of Political Economy 81, 637-654.

Breitung, J., 2002. Nonoparametric Tests for Unit Roots and Cointegration. Journal of Econometrics, 108, 343–364.

Cai, Z., 2007. Trending Time-varying Coefficient Time Series Models with Serially correlated Errors, Journal of Econometrics, 136, 163-188.

Campbell, J., Lo, A., Mackinlay, A., 1997. The Econometrics of Financial Markets, Princeton University Press.

Cavaliere, G., 2004, Unit Root Tests under Time-Varying Variances. Econometric Reviews, Vol. 23, No. 3, 259-292.

Chapman, D., Pearson, N., 2000. Is the Short Rate Drift Actually Nonlinear? Journal of Finance, 55, 355-388.

Chen, X., Hansen, L., Carrasco, M., 2008. Nonlinearity and Temporal Dependence, Cowles Foundation Discussion Paper, No. 1652.

Cox, J., Ingersoll, J., Ross, S., 1985. A Theory of the Term Structure of Interest Rates. Econometrica, Vol. 53, No. 2, 385-407.

Dahlhaus, R., 1996. On the Kullback-Leiber Information divergence of locally stationary processes, Stochastic Processes and their Applications, 62, 139-168.

Dahlhaus, R., 1997. Fitting time series models to nonstationary processes. The Annals of Statistics, Vol.25, No. 1, 1-37.

Dahlhaus, R., Polonik, W., 2009. Empirical Spectral Processes for Locally Stationary Time Series. Bernoulli, Vol.15, No. 1, 1-39.

Dahlhaus, R., Subba Rao, S., 2006. Statistical Inference for Time Varying Arch Processes. The Annals of Statistics, Vol.34, No. 3, 1075-1114.

Dai, Q., Singleton, K., 2000. Specification Analysis of Affine Term Structure Models. Journal of Finance LV, 1943-1978.

Fan, J. Jiang, J. Zhang, C., Zhou, Z., 2003. Time-Dependent Diffusion Models For Term Structure Dynamics. Statistica Sinica, 13, 965-992.

Fryzlewicz, P. Sapatinas, T. Subba Rao, S., 2008, Normalised Least Squares Estimation in Time-Varying ARCH Models. Annals of Statistics Vol. 36, 742-786.

Hansen, B., 2008. Uniform Convergence Rates for Kernel Estimation with Dependent Data. Econometric Theory 24, 726-748.

Härdle, W., Linton, O., 1994. Applied Nonparametric Methods. Cowles Foundation Discussion Paper 1069, The Handbook of Econometrics, Vol. IV.

Hong, Y., Li, H., 2005. Nonparametric Specification Testing for Continuous-Time Models with Application to Term Structure of Interest Rates, The Review of Financial Studies, Vol. 18, No. 1, 37-83.

Hull, J., White, A., 1990. Pricing Interest Rate Derivative Securities. The Review of Financial Studies, Vol.3, No.4. 573-592.

Jiang, G., Knight, J., 1997. A Nonparmetric Approach to the Estimation of Diffusion Processes, with an Application to a Short-Term Interest Rate Model. Econometric Theory, Vol. 13, No. 5, 615-645.

Karatzas, I., Shreve, S., 2000. Brownian Motion and Stochastic Calculus, 2nd E. Springer.

Kloeden, P., Platen, E., 1992. Numerical Solution of Stochastic Differential Equation. Springer, Berlin.

Kristensen, D., 2004. Estimation in Two Classes of Semiparametric Diffusion Models, FMG Discussion Papers DP500. London School of Economics.

Kristensen, D., 2008. Estimation of Partial Differential Equations with Applications in Finance. Journal of Econometrics, 144, 392-408.

Kristensen, D., 2009, Uniform Convergence Rates of Kernel Estimators with Heterogeneous Dependent Data. Econometric Theory, Vol. 25, 1433-1445.

Kim, W., 1999. Nonparametric Analyses of Evolutionary Time Series and Nonlinear Additive ARCH Models. PhD. dissertation, Yale University.

Masry, E., 1996, Multivariate Local Polynomial Regression for Time Series : Uniform Strong Consistency and Rates. Journal of Time Series Analysis, vol 17, 571-599.

Merton, C., 1973. Theory of Rational option Pricing. Bell Journal of Economics and Management Science 4, 141-183.

Meyn, S., Tweedie, R. 1993. Stability of Markovian Processes III: Foster-Lyapunov Criteria for Continuous-Time Processes, Advances in Applied Probability, Vol. 25, No. 3, 518-548.

Müller, H., Stadtmüller, U. 1987. Variable Bandwidth Kernel Estimators of Regression Curves. The Annals of Statistics, Vol. 15, No. 1, 182-201.

Orbe, S., Ferreira, E., Rodríguez-Póo 2000. A Nonaparametric Method to Estimate Time Varying Coefficients under Seasonal Constraints. Journal of Nonparametric Statistics, 12. 779-806.

Phillips, P. C. B., 1973. The Problem of Identification in Finite Parameter Continuous Time Models. Journal of Econometrics 1, 351-362.

Robinson, P., 1983. Nonparametric Estimators For Time Series, Journal of Time Series Analysis Vol. 4 No. 3, 185-207.

Robinson, P., 1989. Nonparametric Estimation of Time-Varying Parameters. Statistical Analysis and Forecasting of Economic Structural Change. Springer-Verlag, 253-264.

Stanton, R. 1997. A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. Journal of Finance 52, 1973-2002.

Appendix A. Proofs of Propositions and Theorems

Condition of Locally Stationary Processes. Consider the following Itô process $X = \{X_t, t_0 \leq t \leq T\}$ and $\tilde{X} = \left\{\tilde{X}_{u,t}, t_0 \leq t \leq T\right\}$.

$$dX_t = \mu(t/T, X_t)dt + \sigma(t/T, X_t)dW_t, \quad X_{t_0} = x_0$$

$$d\tilde{X}_{u,t} = \mu(u, \tilde{X}_{u,t})dt + \sigma(u, \tilde{X}_{u,t})dW_t, \quad \tilde{X}_{u,t_0} = x_0.$$

We will show the following condition holds.

$$\Pr\left\{\max_{1\leq t\leq T}\left|X_{t,T}(\omega) - \tilde{X}_{t/T,t}(\omega)\right| \leq D_T T^{-1/2}\right\} = 1.$$

Euler Approximation associated with the underlying DGP,

$$Y_t = Y_{t-\Delta} + \mu\left(\tau_{t-\Delta}, Y_{t-\Delta}\right)\left(t - (t-\Delta)\right) + \sigma\left(\tau_{t-\Delta}, Y_{t-\Delta}\right)\varepsilon_t,\ Y_0 = x_0,$$

where $\varepsilon_t = W_t - W_{t-\Delta} \sim N(0, \Delta)$, $t = 1, 2, \ldots, T$ and $0 = \tau_0 < \tau_1 < ... < \tau_t = t/T < ... < \tau_T = 1$.

Also, for the $\left\{\tilde{X}_{u,t}\right\}$, for $u \in [0, 1]$,

$$\tilde{Y}_t(u) = \tilde{Y}_{t-\Delta}(u) + \mu\left(u, \tilde{Y}_{t-\Delta}(u)\right)\Delta + \sigma\left(u, \tilde{Y}_{t-\Delta}(u)\right)\varepsilon_t,\ Y_{u,0} = x_0.$$

Therefore, for the $\left\{\tilde{X}_{t/T,t}\right\}$

$$\tilde{Y}_t(t/T) = \tilde{Y}_{t-\Delta}(t/T) + \mu\left(t/T, \tilde{Y}_{t-\Delta}(t/T)\right)\Delta + \sigma\left(t/T, \tilde{Y}_{t-\Delta}(t/T)\right)\varepsilon_t$$

$$\left|X_t - \tilde{X}_{t/T,t}\right| \leq \left|X_t - Y_{\tau_t}\right| + \left|\tilde{X}_{t/T,t} - \tilde{Y}_{\tau_t}(t/T)\right| + \left|Y_{\tau_t} - \tilde{Y}_{\tau_t}(t/T)\right|.$$

Note that it is well known that the euler approximation holds with negligible error. See Kloeden and Platen (1992). We focus on the third term only. Note that our drift term is parameterized as a mean reverting process locally in time. Also, since the diffusion function is unrestricted and therefore, it is difficult to show the following relationship with exact convergence rate. Rather, we provide conservative rate of convergence. In this regard, we restrict ourselves to a certain class of diffusion processes by parameterizing the diffusion function in order to show our definition is not void. It is worth mentioning the following model contains very general affine diffusion processes in Dai and Singleton (2000) among many others.

$$\begin{aligned}
dX_t &= \left(a(t/T) + b(t/T)X_t\right)dt + \left(c(t/T) + d(t/T)X_t\right)dW_t \\
Y_t &= Y_{t-\Delta} + \left(a(\tau_{t-\Delta}) + b(\tau_{t-\Delta})Y_{t-\Delta}\right)\Delta + \left(c(\tau_{t-\Delta}) + d(\tau_{t-\Delta})Y_{t-\Delta}\right)\varepsilon_t \\
\tilde{Y}_t(t/T) &= \tilde{Y}_{t-\Delta}(t/T) + \left(a(t/T) + b(t/T)Y_{t-\Delta}(t/T)\right)\Delta + \left(c(t/T) + d(t/T)Y_{t-\Delta}(t/T)\right)\varepsilon_t.
\end{aligned}$$

We shall proceed by induction. For $t = 1$, due to Assumption 2,

$$\begin{aligned}
&\left|Y_1 - \tilde{Y}_1(\tau_1)\right| \\
&= |Y_0 - \tilde{Y}_0(\tau_1) + (a(\tau_0) - a(\tau_1))\Delta \\
&+ \left(b(\tau_0)Y_0 - b(\tau_1)Y_0 + b(\tau_1)Y_0 - b(\tau_1)\tilde{Y}_0(\tau_1)\right)\Delta \\
&+ (c(\tau_0) - c(\tau_1))\varepsilon_t + \left(d(\tau_0)Y_0 - d(\tau_1)Y_0 + d(\tau_1)Y_0 - d(\tau_1)\tilde{Y}_0(\tau_1)\right)\varepsilon_t| \\
&= O_p\left(T^{-1}\right).
\end{aligned} \tag{27}$$

Therefore, condition (7) holds .

For $t > 1$, suppose the condition holds for $t - \Delta$. Then,

$$\begin{aligned}
&\left|Y_t - \tilde{Y}_t(\tau_t)\right| \\
&= |Y_{t-\Delta} - \tilde{Y}_{t-\Delta}(\tau_t) + (a(\tau_{t-\Delta}) - a(\tau_t))\Delta \\
&+ \left(b(\tau_{t-\Delta})Y_{t-\Delta} - b(\tau_t)Y_{t-\Delta} + b(\tau_t)Y_{t-\Delta} - b(\tau_t)\tilde{Y}_{t-\Delta}(\tau_t)\right)\Delta \\
&+ (c(\tau_{t-\Delta}) - c(\tau_t))\varepsilon_t \\
&+ \left(d(\tau_{t-\Delta})Y_{t-\Delta} - d(\tau_t)Y_{t-\Delta} + d(\tau_t)Y_{t-\Delta} - d(\tau_t)\tilde{Y}_{t-\Delta}(\tau_t)\right)\varepsilon_t| \\
&\leq \left|Y_{t-\Delta} - \tilde{Y}_{t-\Delta}(\tau_t)\right| + |a(\tau_{t-\Delta}) - a(\tau_t)|\Delta \\
&+ \left|(b(\tau_{t-\Delta}) - b(\tau_t))Y_{\tau_{t-1}}\right|\Delta + \left|b(\tau_t)\left(Y_{t-\Delta} - \tilde{Y}_{t-\Delta}(\tau_t)\right)\right|\Delta \\
&+ |c(\tau_{t-\Delta}) - c(\tau_t)||\varepsilon_t| \\
&+ \left|(d(\tau_{t-\Delta}) - d(\tau_t))Y_{\tau_{t-1}}\right||\varepsilon_t| + \left|d(\tau_t)\left(Y_{t-\Delta} - \tilde{Y}_{t-\Delta}(\tau_t)\right)\right||\varepsilon_t| \\
&= O_p\left(T^{-1}\right),
\end{aligned}$$

due to Assumption 2 and (27). In addition, let $M_t = \left|X_{t,T} - \tilde{X}_t(t/T)\right|$ for $t = 1, \ldots, T$ and $M_{\max}$ be $\max\{M_1, \ldots, M_T\}$. For $\varepsilon = O\left(T^{-1/2}\right) > 0$,

$$\begin{aligned}
P(M_{\max} \leq \varepsilon) &= P(M_1 \leq \varepsilon, M_2 \leq \varepsilon, \ldots, M_T \leq \varepsilon) \\
&\geq \prod_{t=1}^{T} P(M_t \leq \varepsilon) \\
&= \exp\left(\sum_{t=1}^{T} \log(1 - (1 - P(M_t \leq \varepsilon)))\right) \\
&\geq \exp\left(-\sum_{t=1}^{T} (1 - P(M_t \leq \varepsilon))\right) \to 1 \text{ as } T \to \infty,
\end{aligned}$$

since $M_t = O_p\left(T^{-1}\right)$ for $\forall t$. This implies (7). ■

Proof of Proposition 1. Due to the triangle inequality,

$$\left|X_t - \tilde{X}_{u,t}\right| \leq \left|X_t - \tilde{X}_{t/T,t}\right| + \left|\tilde{X}_{t/T,t} - \tilde{X}_{u,t}\right|.$$

From the definition of local stationarity, $\left|X_t - \tilde{X}_{t/T,t}\right| = O_p\left(T^{-1/2}\right)$. Therefore, we focus on $\left|\tilde{X}_{t/T,t} - \tilde{X}_{u,t}\right|$. Using Talyor expansion, for t/T in a neighborhood of u_0,

$$\tilde{X}_{t/T,t} = \tilde{X}_{u_0,t} + \left[\frac{t}{T} - u_0\right] \frac{\partial \tilde{X}_{u,t}}{\partial u}\Bigg]_{u=u_0} + \frac{1}{2}\left[\frac{t}{T} - u\right]^2 \frac{\partial^2 \tilde{X}_{u,t}}{\partial u^2}\Bigg]_{u=u_0} + o_p\left(\left[\frac{t}{T} - u\right]^2\right).$$

Since μ, $\sigma \in \mathbb{C}_2\left(C_0\right)$,

$$\left|\tilde{X}_{t/T,t} - \tilde{X}_{u,t}\right| \leq O_p\left(\left|\frac{t}{T} - u\right|\right).$$

Therefore, Proposition 1 follows. ■

Proof of Proposition 2. Due to the triangle inequality,

$$\left|f_u\left(x\right) - g_t\left(x\right)\right| \leq \left|f_u\left(x\right) - f_{t/T}\left(x\right)\right| + \left|f_{t/T}\left(x\right) - g_t\left(x\right)\right|.$$

Let's start with the second term of the right hand side, $\left|f_{t/T}\left(x\right) - g_t\left(x\right)\right|$. For the sake of expositional simplicity, we consider densities g and f for two random variables X and $\tilde{X}$ instead of two stochastic processes. Suppose that

$$|X - \tilde{X}| \leq \epsilon.$$

Then,

$$F_X(x) = \Pr\left[X \leq x\right] = \Pr\left[\tilde{X} + X - \tilde{X} \leq x\right] \leq \Pr\left[\tilde{X} \leq x + \epsilon\right] = F_{\tilde{X}}(x+\epsilon).$$

Likewise, it can be shown that $F_X(x) \geq F_{\tilde{X}}(x-\epsilon)$. By Taylor expansion, we have

$$F_{\tilde{X}}(x) - \epsilon f_{\tilde{X}}(x) + O(\epsilon^2) \leq F_X(x) \leq F_{\tilde{X}}(x) + \epsilon f_{\tilde{X}}(x) + O(\epsilon^2)$$

Since,

$$g_X(x) = \lim_{\delta \to 0} \frac{F_X(x+\delta) - F_X(x)}{\delta}$$

and

$$\begin{aligned}\frac{F_X(x+\delta) - F_X(x)}{\delta} &\leq \frac{F_Y(x+\epsilon+\delta) - F_Y(x+\epsilon)}{\delta} \\ &= f_Y(x) + \epsilon f_Y'(x) + O(\epsilon^2) + O(\delta).\end{aligned}$$

So, under smoothness conditions on the density of the approximating process, we have $f_X(x)$ exists and satisfies

$$f_X(x) = f_Y(x) + \epsilon f_Y'(x) + O(\epsilon^2)$$

as $\epsilon \to 0$. The argument is more complicated when X and $\tilde{X}$ are stochastic processes but is essentially the same. Therefore, due to the definition of local stationarity, with suitably chosen ϵ and δ, $\left|f_{t/T}(x) - g_t(x)\right| = o\left(T^{-2/5}\right)$. With respect to $\left|f_u(x) - f_{t/T}(x)\right|$, due to differentiability and talyor expansion, Proposition 2 follows. ■

Proof of Theorem 1. To begin with, Lemmas we use for the proofs in this section are provided in Appendix B. We begin with the consistency result of $\hat{f}(u,x)$, (18). As a result of Lemma 2, (18) in Theorem 1 follows. Once we have the consistency of the kernel estimator of marginal density function, we proceed to the consistency of the estimators of the drift, (19) and the diffusion functions, (20). For the proof of the statement (19), there is a strand of literature focusing on time varying coefficients. (See Robinson (1988), Orbe *et al* (2005) and Cai (2007)). We adjust their proofs to our setting. Recall

$$\hat{\vartheta}(u) = \left[\sum_{t=1}^{T} K_{ut} Z_t Z_t^{\top}\right]^{-1} \left[\sum_{t=1}^{T} K_{ut} Z_t Y_t\right]$$
$$\tilde{\vartheta}(u) = \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Z}_{ut}^{\top}\right]^{-1} \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Y}_{ut}\right].$$

Due to Proposition 1 with α-mixing $\{X_t\}$, it is shown that in a neighbourhood of u, both $\frac{1}{Th}\sum_{t=1}^{T} K_{ut} Z_t Z_t^{\top}$ and $\frac{1}{Th}\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Z}_{ut}^{\top}$ converge to $M_u := E\left[\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\right]$ where M_u denotes expectation of $\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}$ in the vicinity of the fixed time u. Also, both $\frac{1}{Th}\sum_{t=1}^{T} K_{ut} Z_t Y_t$ and $\frac{1}{Th}\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Y}_{ut}$ converge to $E\left(\tilde{Z}_{ut}\tilde{Y}_{ut}\right)$ by the same token. (See also Lemma A.5 in Dahlhaus and Subba Rao (2006) and Lemma 2 of A.1 in Fryzlewicz *et al.* (2008) for more details.) Meanwhile, let $\tilde{Y}_{ut} - \tilde{Z}_{ut}\vartheta(t/T) = v_{ut}$, then

$$\begin{aligned}
\tilde{\vartheta}(u) - \vartheta(u) &= \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Z}_{ut}^{\top}\right]^{-1} \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Y}_{ut}\right] - \vartheta(u) \\
&= \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Z}_{ut}^{\top}\right]^{-1} \sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \left[\tilde{Z}_{ut}^{\top}\vartheta\left(\frac{t}{T}\right) + v_{ut}\right] - \vartheta(u) \\
&= \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Z}_{ut}^{\top}\right]^{-1} \sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \left\{\tilde{Z}_{ut}^{\top}\left(\vartheta\left(\frac{t}{T}\right) - \vartheta(u)\right)\right\} \\
&+ \left[\sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} \tilde{Z}_{ut}^{\top}\right]^{-1} \sum_{t=1}^{T} K_{ut} \tilde{Z}_{ut} v_{ut} \\
&= \left[M_T(u)\right]^{-1}\left[I_{1T}(u) + I_{2T}(u)\right].
\end{aligned} \tag{28}$$

Also, for $B>0$, let $\sum' = \sum_{|uT-t|\le BTh}$ and $\sum'' = \sum_{|uT-t|>BTh}$.

$$
\begin{aligned}
&E\left[\left\|\frac{1}{Th}\sum K_{ut}\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\left\{\tilde{\vartheta}\left(\frac{t}{T}\right)-\vartheta\left(u\right)\right\}\right\|\right] \\
&\le \sup_{|Tu-t|\le Bh}\left\|\tilde{\vartheta}\left(\frac{t}{T}\right)-\vartheta\left(u\right)\right\|\frac{1}{Th}\sum{}'|K_{ut}|\mathrm{tr}\left(M_u\right)+\frac{C}{Th}\sum{}''|K_{ut}|tr\left(M\right) \\
&\le C\left[\sup_{|Tu-t|\le Bh}\left\|\tilde{\vartheta}\left(\frac{t}{T}\right)-\vartheta\left(u\right)\right\|+\int_{|u|\ge B}|K(u)|du\right]+o(1).
\end{aligned}
$$

Finally,

$$
E\left[\left\|\frac{1}{Th}\sum K_{ut}\tilde{Z}_{ut}v_t\right\|\right]=\frac{1}{(Th)^2}\sum K_{ut}^2\mathrm{tr}\left(M\right)\sigma^2=O_p\left(\frac{1}{Th}\right).
$$

Therefore, $\tilde{\vartheta}\left(u\right)\xrightarrow{p}\vartheta\left(u\right)$, which implies $\hat{\vartheta}\left(u\right)\xrightarrow{p}\vartheta\left(u\right)$. Now, Slutsky theorem yields the consistency result of $\hat{\theta}$, given $\hat{\alpha}=-\dfrac{\hat{b}}{\hat{a}}$, $\hat{\beta}=-\ln(\hat{b}+1)$. Hence, the statement (19) follows. For the consistency result of $\hat{\sigma}^2$, (20), recall that

$$
\begin{aligned}
\hat{\sigma}^2(u,x)-\sigma_0^2(u,x) &= \frac{2}{\hat{f}(u,x)}\int_{-\infty}^{x}\hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy-\frac{2}{f_0(u,x)}\int_{-\infty}^{x}\mu(u,y;\theta_0)f_0(u,y)dy \\
&= \frac{2}{\hat{f}(u,x)}\int_{-\infty}^{x}\mu(u,y;\theta_0)f_0(u,y)dy-\frac{2}{\hat{f}(u,x)}\int_{-\infty}^{x}\mu(u,y;\theta_0)f_0(u,y)dy \\
&+\frac{2}{\hat{f}(u,x)}\int_{-\infty}^{x}\hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy-\frac{2}{f_0(u,x)}\int_{-\infty}^{x}\mu(u,y;\theta_0)f_0(u,y)dy \\
&= 2\int_{-\infty}^{x}\mu(u,y;\theta)f_0(u,y)dy\left[\frac{1}{\hat{f}(u,x)}-\frac{1}{f_0(u,x)}\right] \\
&+\frac{2}{\hat{f}(u,x)}\left[\int_{-\infty}^{x}\hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy-\int_{-\infty}^{x}\mu(u,y;\theta_0)f_0(u,y)dy\right] \\
&= I_1+I_2.
\end{aligned}
$$

For I_1,

$$
I_1=2\int_{-\infty}^{x}\mu(u,y;\theta)f_0(u,y)dy\left[\frac{1}{\hat{f}(u,x)}-\frac{1}{f_0(u,x)}\right].
$$

Let $H:=\left[\dfrac{1}{\hat{f}(u,x)}-\dfrac{1}{f_0(u,x)}\right]$.

$$
|H|=-\frac{1}{\hat{f}\left(u,x\right)f_0\left(u,x\right)}\left(\hat{f}\left(u,x\right)-f_0\left(u,x\right)\right)+o_p\left(||\hat{f}\left(u,x\right)-f_0\left(u,x\right)||\right).
$$

Given that $f_u\left(x\right)$ is bounded away from zero and $\hat{f}\left(u,x\right)\xrightarrow{p}f_0\left(u,x\right)$, $|H|\xrightarrow{p}0$, which implies $|I_1|\xrightarrow{p}0$. On the other hand, it can be easily shown $|I_2|\xrightarrow{p}0$ given the consistency of the marginal density and the drift shown before. Collecting the results, the statement (20) follows. ■

Proof of Theorem 2. Given the relationship between $\hat{\theta}$ and $\hat{\vartheta}$, we start with the asymptotic normality of $\hat{\vartheta}$. With Lemma 3, the asymptotic normality property of $\hat{\mu}$ follows as soon as one can notice that $\sigma(u, X_t)\varepsilon_t$ are martingale differences with the assumption of existence of its finite $(2+\kappa)$th moment where $\kappa > 0$. The statement (21) in Theorem 2 follows. For the part with respect to $\hat{\sigma}^2$, we start with the asymptotic normality of our kernel density estimator. Due to the asymptotic normality of kernel estimator of marginal density functions (Lemma 4), we proceed to prove the asymptotic normality of the diffusion function. Let σ_0^2 be the true functional of the diffusion term. Note that the diffusion consists of the drift and the marginal density. The drift $\hat{\mu}$ are estimated at the relatively faster rate than kernel density estimator $\hat{f}_u(x)$. Moreover, it is a part of integrand of the diffusion function. Consequently, the asymptotic distribution of $\hat{\sigma}^2$ is affected only by the asymptotic distribution of $\hat{f}_u(x)$, not by that of $\hat{\mu}$. Given the asymptotic distribution of $\hat{f}(u, x)$,

$$
\begin{aligned}
\hat{\sigma}^2(u,x) - \sigma_0^2(u,x) &= \frac{2}{\hat{f}(u,x)} \int_{-\infty}^{x} \hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy - \frac{2}{f_0(u,x)} \int_{-\infty}^{x} \mu(u,y;\theta_0) f_0(u,y)dy \\
&= \frac{2}{\hat{f}(u,x)} \int_{-\infty}^{x} \mu(u,y;\theta_0) f_0(u,y)dy - \frac{2}{\hat{f}(u,x)} \int_{-\infty}^{x} \mu(u,y;\theta_0) f_0(u,y)dy \\
&+ \frac{2}{\hat{f}(u,x)} \int_{-\infty}^{x} \hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy - \frac{2}{f_0(u,x)} \int_{-\infty}^{x} \mu(u,y;\theta_0) f_0(u,y)dy \\
&= 2 \int_{-\infty}^{x} \mu(u,y;\theta) f_0(u,y)dy \left[\frac{1}{\hat{f}(u,x)} - \frac{1}{f_0(u,x)} \right] \\
&+ \frac{2}{\hat{f}(u,x)} \left[\int_{-\infty}^{x} \hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy - \int_{-\infty}^{x} \mu(u,y;\theta_0) f_0(u,y)dy \right] \\
&= I_1 + I_2.
\end{aligned}
$$

For I_1,

$$
I_1 = 2 \int_{-\infty}^{x} \mu(u,y;\theta) f_0(u,y)dy \left[\frac{1}{\hat{f}(u,x)} - \frac{1}{f_0(u,x)} \right],
$$

while

$$
\left[\frac{1}{\hat{f}(u,x)} - \frac{1}{f_0(u,x)} \right] = \frac{-(\hat{f}(u,x) - f_0(u,x))}{\hat{f}(u,x) f_0(u,x)}.
$$

For I_2,

$$
I_2 = \frac{2}{\hat{f}(u,x)} \left[\int_{-\infty}^{x} \left[\hat{\mu}(u,y;\hat{\theta})\hat{f}(u,y)dy - \mu(u,y;\theta_0) f_0(u,y) \right] dy \right].
$$

Since I_2 consists of integration of estimate and the integrand has the same convergence rate as $\hat{f}\,(u,x)$, it can be easily shown that I_2 is smoother and hence smaller order than I_1. Therefore,

$$
I_2 = o\left((Th_1h_2)^{-1/2} \right).
$$

Therefore, due to Lemma 4 and continuous mapping theorem, The statement (22) in Theorem 2 follows. Moreover, due to Robinson (1983), for distinct time points u, x and u^*, x^*, we have $\hat{\theta}(u)$, $\hat{\sigma}^2(u,x)$ are asymptotically independent of $\hat{\theta}(u^*)$, $\hat{\sigma}^2(u^*,x^*)$ respectively. Collecting the results thus far completes the proof of Theorem 2. ■

Proof of Theorem 3. Let's start with the uniform convergence of the drift function. Because we assume X_t is strongly mixing with geometrically decreasing mixing coefficients and u can be considered from $U[0,1]$ with exponential decay. It could be considered as the straighforward extension of Kristensen (2009). We introduce the following notation for our argument. For our triangular array of interest,

$$
\begin{aligned}
B_0 &= \sup_{t,T} \sup_{x\in\mathbb{R}} g_{t,T}(x) \\
B_{Y,1} &= \sup_{t,T} \sup_{x\in\mathbb{R}} E\left[Y_{t,T}|X_{t,T}=x\right] g_{t,T}(x) \\
B_{Y,2} &= \sup_{T} \sup_{|t-j|\geq M} \sup_{x\in\mathbb{R}} E\left[Y_{t,T}Y_{j,T}|X_{t,T}=x,\ X_{j,T}=y\right] g_{t,j,T}(x,y),
\end{aligned}
$$

where $g_{t,T}(x)$ is the density of $\{X_{t,T}\}$ and M is some positive number and $\sup_{t,T} = \sup_{T\geq 1}\sup_{1\leq t\leq T}$. Under Assumptions 1 - 8, conditions for Theorem 1 in Kristensen (2009) are met and application of Theorem 1 in Kristensen can be used.

Recall that the proposed estimator of the drift function is

$$
\hat{\vartheta}(u) = \left[\sum K_{ut} Z_t Z_t^{\top}\right]^{-1}\left[\sum K_{ut} Z_t Y_t\right].
$$

We consider a and b separately. For example, let's start with $\hat{b}$. Let $\frac{1}{Th}\sum K_{ut}\frac{1}{Th}\sum K_{ut}X_t^2 - \left(\frac{1}{Th}\sum K_{ut}X_t\right)^2 = \hat{\Phi}_d^b$ and $\frac{1}{Th}\sum K_{ut}\frac{1}{Th}\sum K_{ut}X_tY_t - \frac{1}{Th}\sum K_{ut}X_t\frac{1}{Th}\sum K_{ut}Y_t = \hat{\Phi}_n^b$. Note that $\hat{\Phi}_d^b$ and $\hat{\Phi}_n^b$ are local demeaned values. All assumptions provided in Theorem 1 of Kristensen (2009) are met with $Y_{t,T} = X_{t,T}X_{t-1,T}$ and $Y_{t,T} = X_{t,T}^2$. In particular, B_0, $B_{Y,1}$, $B_{Y,2}$ are all bounded. Due to Kristensen (2009), we obtain:

$$
\begin{aligned}
\sup_{u\in[\delta_T,1-\delta_T]} |\hat{\Phi}_d^b - E\left[\hat{\Phi}_d^b\right]| &= O_p\left(r_*(T)\right) \\
\sup_{u\in[\delta_T,1-\delta_T]} |\hat{\Phi}_n^b - E\left[\hat{\Phi}_n^b\right]| &= O_p\left(r_*(T)\right).
\end{aligned}
$$

where $r_*(T) := \left(\frac{\ln T}{Th}\right)^{1/2}$. Also, due to standard nonparametric estimation manipulation,

$$
\begin{aligned}
\sup_{u\in[\delta_T,1-\delta_T]} |E\left[\hat{\Phi}_d^b\right] - \Phi_d^b| &= O_p\left(h^2\right) \\
\sup_{u\in[\delta_T,1-\delta_T]} |E\left[\hat{\Phi}_n^b\right] - \Phi_n^b| &= O_p\left(h^2\right).
\end{aligned}
$$

Therefore,

$$\sup_{u\in[\delta_T,1-\delta_T]} |\hat{\Phi}_d^b - \Phi_d^b| = O_p\left(r_*\left(T\right)\right) + O\left(h^2\right)$$
$$\sup_{u\in[\delta_T,1-\delta_T]} |\hat{\Phi}_n^b - \Phi_n^b| = O_p\left(r_*\left(T\right)\right) + O\left(h^2\right).$$

By the Taylor expansion,

$$\begin{aligned}\left|\hat{b}\left(u\right) - b\left(u\right)\right| &= \left|\frac{\hat{\Phi}_n^b}{\hat{\Phi}_d^b} - b\left(u\right)\right| = \left|\frac{\hat{\Phi}_n^b}{\hat{\Phi}_d^b} - \frac{\Phi_n^b}{\Phi_d^b}\right| \\ &\leq \frac{|\hat{\Phi}_n^b - \Phi_n^b|}{\Phi_d^b} + \frac{|\Phi_n^b|}{\left(\Phi_d^b\right)^2}|\hat{\Phi}_d^b - \Phi_d^b|.\end{aligned}$$

Due to the same argument as in Kristensen (2009), the uniform convergence rate of the above equation is determined by $|\hat{\Phi}_n^b - \Phi_n^b|$ and $|\hat{\Phi}_d^b - \Phi_d^b|$. Therefore,

$$\sup_{u\in[\delta_T,1-\delta_T]} |\hat{b}\left(u\right) - b\left(u\right)| = O_p\left(r_*\left(T\right)\right) + O\left(h^2\right).$$

Also, as for $\hat{a}$, the method in Hansen (2008) applies again with the previous result for $\hat{b}$ and the rate of uniform convergence of $\hat{a}$ is determined by that of $\hat{b}$. Since $\theta = [\alpha, \beta]^\top$ is obtained from the relation, $\alpha = -\frac{a}{b}$ and $\beta = -ln(b+1)$, the result follows. For the uniform convergence of the diffusion function, notice that the uniform convergence of the proposed estimator of the diffusion function is determined by that of the proposed estimator of the density function. Therefore, to provide the uniform consistency result of $\hat{\sigma}^2$, we start with the uniform consistency and rates for $\hat{f}(x,u)$. Due to Lemma 5, the result follows. ■

Appendix B. Lemmas

Lemma 1 (Benedetti, 1977) *Suppose K is continuous, and is such that $K(u)$ is nonincreasing for $u > 0$, and nondecreasing for $u < 0$, and suppose $\int K^r(u)du < \infty$. If there exists some $\bar{\Delta}$ such that $\Delta/n \geq \max(x_i - x_{i-1})$ for all n, where $x_0 = 0, x_{n+1} = 1$, and if $nh_n \to \infty$, then for $x \in (0,1)$*

$$h_n^{-1}\sum_{i=1}^{n}(x_i - x_{i-1})K^r\left(\frac{x - x_i}{h_n}\right) = \int K^r(u)du + O\left((nh_n)^{-1}\right). \tag{29}$$

Proof of Lemma 1. See Orbe *et al* (2000). ■

Lemma 2 *Under the Assumptions 2, 3 and 4, then*

$$E\left[\hat{f}(u,x)-g_0(u,x)\right]^2 = \left[\left(\frac{h_2^2 f_0^{(2)}(u,x)+h_1^2\ddot{f}_0\left(u,x\right)}{2}\right)\mu_2\left(K\right)+o\left(h_1^2+h_2^2+T^{-1/2}\right)\right]^2 + \frac{f_0\left(u,x\right)}{Th_1h_2}||K||_2^4+o((Th_1h_2)^{-1}).$$

Proof of Lemma 2. The terms in square bracket of $E\left[\hat{f}(u,x)-g_0(u,x)\right]^2$ come from the bias of $\hat{f}(u,x)$, while the last two terms come from variance of $\hat{f}(u,x)$. Note that as is the case with standard nonparametric density estimation methods, our density estimator is subject to some bias and shares most of the basic statistical properties with the standard ones in the nonparametric literature. Note that if the underlying process is stationary, then $\ddot{f}_0\left(u,x\right)=0$.The proof is built upon the approach of Bosq (1998). We modify it slightly according to our locally stationary case. Let's begin with Bias term of $\hat{f}(u,x)$. Following Lemma 1 by slight modification,

$$h^{-1}\sum_{t=1}^{T}(\frac{t}{T}-\frac{t-1}{T})K^r\left(\frac{x-t/T}{h}\right)=(Th)^{-1}\sum_{t=1}^{T}K^r\left(\frac{x-t/T}{h}\right)=\int K^r(u)du+O\left((Th)^{-1}\right).$$

Using the standard methods of kernel estimators combined with the assumption of local stationarity,

$$E(\hat{f}(u,x))-g_0(u,x)=h_2^2\frac{f_0^{(2)}(u,x)}{2}\mu_2\left(K\right)+o\left(h_2^2\right) + h_1^2\frac{\ddot{f}_0\left(u,x\right)}{2}\mu_2\left(K\right)+o\left(h_1^2\right), \tag{30}$$

due to the assumption of $f(u,x) \in \mathbb{C}_2\left(b\right)$. More specifically, due to Proposition 2,

$$\begin{aligned}
E(\hat{f}_u(x)) - g_u(x) &= E\left(\hat{f}_u\left(x\right)\right) - E\left(\frac{1}{Th_1}\sum K\left(\frac{u-t/T}{h_1}\right)f_u\left(x\right)\right) \\
&+ E\left(\frac{1}{Th_1}\sum K\left(\frac{u-t/T}{h_1}\right)f_u\left(x\right)\right) - g_u(x) \\
&= \frac{1}{Th_1}\sum K\left(\frac{u-t/T}{h_1}\right)\left[\frac{1}{h_2}E\left(K\left(\frac{X_t-x}{h_2}\right)\right) - f_u\left(x\right)\right] \\
&+ E\left(\frac{1}{Th_1}\sum K\left(\frac{u-t/T}{h_1}\right)\left(f_u\left(x\right) - g_u\left(x\right)\right)\right) + O\left(\left(Th_1\right)^{-1}\right) \\
&= \left[1 + O\left(\left(Th_1\right)^{-1}\right)\right]\left[h_2^2\frac{f_u^{(2)}(x)}{2}\int w^2K\left(w\right)dw + o\left(h_2^2\right)\right] \\
&+ E\left(\frac{1}{Th_1}\sum K\left(\frac{u-t/T}{h_1}\right)\left|\frac{t}{T}-u\right|\dot{f}_0\left(u,x\right)\right) \\
&+ E\left(\frac{1}{Th_1}\sum K\left(\frac{u-t/T}{h_1}\right)\frac{1}{2}\left|\frac{t}{T}-u\right|^2\ddot{f}_0\left(u,x\right)\right) + o_p\left(\left|\frac{t}{T}-u\right|^2 + T^{-2/5}\right) \\
&= h_2^2\frac{f_0^{(2)}(u,x)}{2}\mu_2\left(K\right) + h_1^2\frac{\ddot{f}_0\left(u,x\right)}{2}\mu_2\left(K\right) + o_p\left(h_1^2 + h_2^2 + T^{-2/5}\right).
\end{aligned}$$

Now, we turn to variance part of Mean Squared Error of $\hat{f}(u,x)$. By definition,

$$\begin{aligned}
\mathrm{var}(\hat{f}(u,x)) &= (Th_1h_2)^{-2}\mathrm{var}\left(\sum_{t=1}^{T}K\left(\frac{u-t/T}{h_1}\right)K\left(\frac{x-X_t}{h_2}\right)\right) \\
&= (Th_1h_2)^{-2}\left[\begin{array}{c}\sum_{t=1}^{T}\mathrm{var}\left(K\left(\frac{u-t/T}{h_1}\right)K\left(\frac{x-X_t}{h_2}\right)\right) + \\ \sum_{s\neq t}\mathrm{cov}\left(K\left(\frac{u-t/T}{h_1}\right)K\left(\frac{x-X_t}{h_2}\right), K\left(\frac{u-s/T}{h_1}\right)K\left(\frac{x-X_s}{h_2}\right)\right)\end{array}\right] \\
&= I_1 + I_2.
\end{aligned}$$

Since we deal with dependent data, we introduce the following local measure of dependence for our asymptotic analysis.

$$\varsigma_{t,s} = f_{(X_{u,t},X_{u,s})}(y,z) - f_{(X_{u,t})}\left(y\right)\otimes f_{(X_{u,t})}\left(z\right). \tag{31}$$

For each couple $\left(t,s\right)$, $t\neq s$, $f_{(X_{t,u},X_{s,u})}(y,z)$ denotes the joint PDF of $(X_{t,u},X_{s,u})$ in the vicinity of a fixed time point u. Also, $\varsigma_{t,s}$ satisfies one of the following conditions:

Condition A-1

1. $\zeta_p = \sup_{|t-s|\geq 1}||\varsigma_{t,s}||_p < \infty$, **for some** $p \in (2,\infty)$

2. $|\varsigma_{t,s}\left(x\right) - \varsigma_{t,s}\left(u\right)| \leq C||x-u||^{\theta}$ **for some constant** C.

First, we start with I_1. Using the standard methods,

$$|I_1| \xrightarrow{p} \frac{f_u\left(x\right)}{Th_1h_2}\left[\int K^2(r)dr\right]\left[\int K^2\left(w\right)dw\right] = O_p((Th_1h_2)^{-1}). \tag{32}$$

The above value goes to zero as $Th_1h_2 \to \infty$ and $\max(h_1, h_2) \to 0$, provided that $T \to \infty$, using the standard Bochner's theorem for i.i.d. case. Meanwhile,

$$
\begin{aligned}
|I_2| &= \left| \frac{2}{(Th_1h_2)^2} \sum_{s=1}^{T-1} \sum_{t=1}^{T-t} K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-s/T}{h_1}\right) \operatorname{cov}\left(K\left(\frac{x-X_t}{h_2}\right), K\left(\frac{x-X_s}{h_2}\right)\right) \right| \\
&\leq 2(Th_1)^{-2} \sum_{s=1}^{T-1} \sum_{t=1}^{T-t} \left| K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-s/T}{h_1}\right) \right| |I_3|, \\
&\text{where } |I_3| = \left| \operatorname{cov}\left(\frac{1}{h_2} K\left(\frac{x-X_{t,u}}{h_2}\right), \frac{1}{h_2} K\left(\frac{x-X_{s,u}}{h_2}\right)\right) \right|.
\end{aligned}
$$

Note that the following relations hold.

$$
\begin{aligned}
|I_3| &\leq \left| \operatorname{cov}\left(\frac{1}{h_2} K\left(\frac{x-X_{t,u}}{h_2}\right), \frac{1}{h_2} K\left(\frac{x-X_{s,u}}{h_2}\right)\right) \right| \\
&= h_2^{-2} \left| E\left\{K\left(\frac{x-X_{t,u}}{h_2}\right) K\left(\frac{x-X_{s,u}}{h_2}\right)\right\} - EK\left(\frac{x-X_{t,u}}{h_2}\right) EK\left(\frac{x-X_{s,u}}{h_2}\right) \right| \\
&\leq h_2^{-2} \int\int \left| K\left(\frac{x-y}{h_2}\right) K\left(\frac{x-z}{h_2}\right) \right| \left| f_u(y,z) - f_u(y) f_u(z) \right| dydz \\
&= h_2^{-2} \int\int \left| K\left(\frac{x-y}{h_2}\right) K\left(\frac{x-z}{h_2}\right) \right| \left| \varsigma_{t,s} \right| dydz,
\end{aligned}
$$

where $\varsigma_{t,s}$ is defined in (31).

In order to proceed further, we need to choose an appropriate bound on the above value by using α-mixing coefficient. For α-mixing (strong mixing), if Condition A–1-1 holds, due to Hölder inequality,

$$
|I_3| \leq \sup_{|s-t|\geq 1} ||\varsigma_{t,s}||_p h_2^{-2/p} ||K||_q^2, \tag{33}
$$

where $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

On the other hand, due to Billingsley's inequality,[8]

$$
|I_3| \leq h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|s-t|}\right|.
$$

Therefore,

$$
|I_3| \leq \min\left(\zeta_p h_2^{-2/p} ||K||_q^2, h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|s-t|}\right|\right).
$$

[8] (Billingsley's inequality) If $Y \in L^\infty(\sigma(X_s, s \leq t))$ and $Z \in L^\infty(\sigma(X_s, s \geq t+k))$, then

$$
|Cov(Y,Z)| \leq 4||Y||_\infty ||X||_\infty \alpha_k
$$

Consequently,

$$
\begin{aligned}
|I_2| &\leq \frac{2}{(Th_1)^{-2}} \sum_{s=1}^{T-1} \sum_{t=1}^{T-t} \left| K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-s/T}{h_1}\right) \right| \\
&\times \left| \min\left(\zeta_p h_2^{-2/p} ||K||_q^2, h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|t-s|}\right|\right) \right| \\
&\leq \frac{2}{(Th_1)^{-2}} \sum_{j=1}^{T-1} \sum_{t=1}^{\lfloor T/2 \rfloor} \left| K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-(t+j)/T}{h_1}\right) \right| \\
&\times \left| \min\left(\zeta_p h_2^{-2/p} ||K||_q^2, h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|j|}\right|\right) \right| \text{ where } j = |s-t| \\
&= I_4 + I_5. \qquad (34)
\end{aligned}
$$

Since K is Lipschitz continuous and has a bounded support,

$$
\frac{1}{(Th_1)^2} \sum_{t=1}^{\lfloor T/2 \rfloor} \left| K\left(\frac{u-t/T}{h_1}\right) \left[K\left(\frac{u-(t+j)/T}{h_1}\right) - K\left(\frac{u-t/T}{h_1}\right) \right] \right| = o\left(\frac{1}{Th_1}\right). \qquad (35)
$$

We can consider two cases:

1. $\min\left(\zeta_p h_2^{-2/p} ||K||_q^2, h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|t-s|}\right|\right) = \zeta_p h_2^{-2/p} ||K||_q^2$, i.e. $1 \leq |s-t| \leq \xi_T$ where $\xi_T \simeq h_2^{-2/q\beta}$, then

$$
|I_4| = \frac{2}{(Th_1)^2} \sum_{j=1}^{\xi_T} \left| \zeta_p h_2^{-2/p} ||K||_q^2 \right| \sum_{t=1}^{\lfloor T/2 \rfloor} \left| K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-(t+j)/T}{h_1}\right) \right|.
$$

From (35)

$$
\begin{aligned}
|I_4| &= \sum_{j=1}^{\xi_T} \left| \zeta_p h_2^{-2/p} ||K||_q^2 \right| \left[\frac{1}{(Th_1)^2} \sum_{t=1}^{\lfloor T/2 \rfloor} K^2\left(\frac{u-t/T}{h_1}\right) + o\left(\frac{1}{Th_1}\right) \right] \\
&\leq \frac{1}{Th_1} \sum_{j=1}^{\xi_T} \left| \zeta_p h_2^{-2/p} ||K||_q^2 \right| \left(\frac{1}{Th_1} \sum_{t=1}^{T} K^2\left(\frac{u-t/T}{h_1}\right) + o(1) \right).
\end{aligned}
$$

For $|I_4|$, $Th_1h_2|I_4| = o\left(h_2^{(\beta(p-2)-2(p-1))/\beta p}\right)$. Since we assume $\beta > \dfrac{2(p-1)}{p-2}$, $Th_1h_2|I_4| = o(1)$.

2. $\min\left(\zeta_p h_2^{-2/p} ||K||_q^2, h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|t-s|}\right|\right) = h_2^{-2} 4 \left(\sup |K(v)|\right)^2 \left|\alpha_{|t-s|}\right|$, i.e. $|s-t| \leq \xi_T$

where $\xi_T \simeq h_2^{-2/q\beta}$, then

$$|I_5| = \frac{2}{(Th_1)^2} \sum_{j>\xi_T} \sum_{t=1}^{\lfloor T/2 \rfloor} \left| K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-(t+j)/T}{h_1}\right) \right| h_2^{-2} 4 \left(\sup|K(v)|\right)^2 \left|\alpha_{|j|}\right|$$

$$\leq \frac{8\left(\sup|K(v)|\right)^2}{T^2 h_1^2 h_2^2} \sum_{j>\xi_T} \gamma j^{-\beta} \sum_{t=1}^{\lfloor T/2 \rfloor} \left| K\left(\frac{u-t/T}{h_1}\right) K\left(\frac{u-(t+j)/T}{h_1}\right) \right|.$$

From (35),

$$|I_5| \leq \frac{const.}{h_2^2} \sum_{j>\xi_T} \gamma j^{-\beta} \left[\frac{1}{(Th_1)^2} \sum_{t=1}^{\lfloor T/2 \rfloor} K^2\left(\frac{u-t/T}{h_1}\right) + o\left(\frac{1}{Th_1}\right) \right]$$

$$\leq \frac{const.}{Th_1h_2} \left[\frac{1}{h_2} \sum_{j>\xi_T} \gamma j^{-\beta} \right] \left(\frac{1}{Th_1} \sum_{t=1}^{T} K^2\left(\frac{u-t/T}{h_1}\right) + o(1) \right).$$

For $|I_5|$, $Th_1h_2|I_5| = o(1)$ .Therefore,

$$Th_1h_2|I_2| = o(1). \tag{36}$$

From (30), (32) and (36), Lemma 2 is proved. If Condition A–1-2 holds, the proof is almost the same except that (33) should be changed as follows.

$$|I_3| \leq C\left[\alpha_{|s-t|}\right]^{1/3}.$$

■

Lemma 3 *Under the Assumption 2 - 6, $\hat{\vartheta}$ is pointwise asymptotically normally distributed as follows.*

$$\sqrt{Th}\left\{\hat{\vartheta}(u) - \vartheta_0(u) - h^2 \text{bias}_{\hat{\theta}}\right\} \quad \xrightarrow{d} \quad N\left(0, ||K||_2^2 M_u^{-1} \Omega_u M_u^{-1}\right) \tag{37}$$

where $M_u := E_u\left[\tilde{Z}_{ut}\tilde{Z}_{ut}^\top\right]$, $\Omega_u := E_u\left(\tilde{Z}_{ut}\tilde{Z}_{ut}^\top \sigma_u^2\left(\tilde{X}_{u,t}\right)\right)$ *and*
$\text{bias}_{\hat{\vartheta}} = [\mu_2(K)] M_u^{-1} E_u\left[\tilde{Z}_{ut}\tilde{Z}_{ut}^\top\left\{\vartheta_0^{(1)}(u)\dot{f}_0\left(u, \tilde{X}_{u,t}\right)/f_0\left(u, \tilde{X}_{u,t}\right) + \frac{1}{2}\vartheta_0^{(2)}(u)\right\}\right]$.

Proof of Lemma 3. Our approach is based on Robinson (1989) and Cai (2007) on top of the previous argument for consistency results of $\hat{\vartheta}(u)$. Continuing from (28),

$$\hat{\vartheta}(u) = \vartheta(u) + [M_T(u)]^{-1}[I_{1T}(u) + I_{2T}(u)]$$

where

$$M_T(u) := \frac{1}{Th}\sum_{t=1}^{T} K_{ut}\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top} \tag{38}$$

$$I_{1T}(u) := \frac{1}{Th}\sum_{t=1}^{T} K_{ut}\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\left(\vartheta\left(\frac{t}{T}\right) - \vartheta(u)\right) \tag{39}$$

$$I_{2T}(u) := \frac{1}{Th}\sum_{t=1}^{T} K_{ut}\tilde{Z}_{ut}v_{ut}. \tag{40}$$

For (38), $M_T(u) = M_u + o(1)$ where $M_u := E_{u_0}\left[\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\right]$ as before. For (39),

$$\begin{aligned}
E\left[I_{1T}(u)\right] &= E\left[K_{ut}\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\left(\vartheta\left(\frac{t}{T}\right) - \vartheta(u)\right)\right] \\
&= \int\int z_1 z_1^{\top}\left(\vartheta(u+hv) - \vartheta(u)\right) f_0(u+hv, x) K(v)\, dx dv \\
&= \int\int z_1 z_1^{\top}\left[hv\vartheta^{(1)}(u) + 1/2h^2v^2\vartheta^{(2)}(u)\right]\left[f_0(u,x) + \dot{f}_0(u,x)hv\right] K(v)\, dx dv \\
&= \int\int z_1 z_1^{\top}\left[\dot{f}_0(u,x)\vartheta^{(1)}(u)h^2v^2 + 1/2 f_0(u,x)\vartheta^{(2)}(u)h^2v^2\right] K(v)\, dx dv + o\left(h^2\right) \\
&= h^2\int v^2 K(v)\, dv \int z_1 z_1^{\top}\left[\frac{\dot{f}_0(u,x)\vartheta^{(1)}(u)}{f_0(u,x)} + 1/2\vartheta^{(2)}(u)\right] f_0(u,x)dx + o\left(h^2\right) \\
&= h_2\left[\mu_2(K)\right] E_u\left[\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\left\{\vartheta_0^{(1)}(u)\dot{f}_0\left(u, \tilde{X}_{u,t}\right)/f_0\left(u, \tilde{X}_{u,t}\right) + \frac{1}{2}\vartheta_0^{(2)}(u)\right\}\right] + o\left(h^2\right).
\end{aligned}$$

For (40), note that $\{v_{ut}\}$ is a martingale difference sequence.

$$\sqrt{Th}I_{2T}(u) \xrightarrow{d} N\left(0, ||K||_2^2\Omega_u\right),$$

where $\Omega_u := E_u\left(\tilde{Z}_{ut}\tilde{Z}_{ut}^{\top}\sigma_u^2\left(\tilde{X}_{u,t}\right)\right)$. Collecting all of the results above, Lemma 3 follows. ■

Lemma 4 *Under the same setting as Theorem 1 with Assumption 5, in the neighborhood of a certain time point u, for $|u - t/T| \to 0$, we have* $\text{bias}_{\hat{f}} = O\left(h_1^2 + h_2^2\right)$ *and*

$$\sqrt{Th_1h_2}\left\{\hat{f}(u,x_i) - g_0(u,x_i) - \text{bias}_{\hat{f}}(u,x_i)\right\}_{i=1}^{k} \xrightarrow{d} N(0, V_f), \tag{41}$$

where $V_f = diag\left\{V_f(u,x_i)\right\},\ i = 1,\ldots,k,\quad V_f(u,x_i) = ||K||_2^4 f(u,x_i)$ *and*

$$\text{bias}_{\hat{f}}(u,x_i) = \left(\frac{h_2^2 f_0^{(2)}(u,x_i) + h_1^2\ddot{f}_0(u,x_i)}{2}\right)\mu_2(K).$$

Moreover, under the conditions of the above Lemma, $V_f(u, x_i)$ is consistently estimated by

$$||K||_2^4 \hat{f}(u, x_i) \text{ for } i = 1, \ldots, k. \tag{42}$$

Proof of Lemma 4.

Our method is a slight modification of Bosq (1998). To begin with, we need to prove that for $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_k) \neq \mathbf{0}$,

$$(Th_1h_2)^{1/2} \sum_{i=1}^{k} \lambda_i \frac{\hat{f}(u, x_i) - E\hat{f}(u, x_i)}{\left(f_0(u, x_i)\{||K||_2^2\}^2\right)^{1/2}} \xrightarrow{d} \sum_{i=1}^{k} \lambda_i N_i, \tag{43}$$

where N_i denotes a standard normal random variable. Let S_T be defined as follows.

$$S_T = \sum_{t=1}^{T} Z_T(t),$$

where

$$Z_T(t) = \sum_{i=1}^{k} \frac{\lambda_i}{\left(f_u(x_i)\,||K||_2^2\right)^{1/2}} \left[K\left(\frac{u - t/T}{h_1}\right) K\left(\frac{x_i - X_t}{h_2}\right) - EK\left(\frac{u - t/T}{h_1}\right) K\left(\frac{x_i - X_t}{h_2}\right)\right],$$

where $t = 1, \ldots, T$. We can consider the following blocks to make use of Bradley's theorem (1983). $V_T(j) = \sum_{i=1}^{p} Z_T((j-1)(p+q)+i)$, $V_T'(j) = \sum_{i=1}^{q} Z_T(jp + (j-1)q + i)$, $\delta_T = \sum_{t=r(p+q)+1}^{T} Z_T(t)$ where $j = 1, \ldots, r$ and $r(p+q) \leq T < r(p+q+1)$. It is obvious that the contribution of δ_T is negligible and thus neglected afterwards when $T \to \infty$. Due to Bradley's theorem, there exist independent random variables $W_T(j)$, $j = 1, \ldots, r$ such that the probability distributions of $W_T(j)$ and $V_T(j)$ are identical and $P(|V_T(j) - W_T(j)| > \xi_T) \leq 18(||V_T(j)||/\xi_T)\,\alpha(q)$, $j = 1, \ldots, r$, with $\xi_T = \varepsilon(rph_1h_2)^{1/2}$ for some $\varepsilon > 0$.

$$\begin{aligned} P\left(\left|\sum_{j=1}^{r} V_T(j)/(rph_1h_2)^{1/2}\right| > \varepsilon\right) &\leq P\left(\left|\sum_{j=1}^{r} (V_T(j) - W_T(j))/(rph_1h_2)^{1/2}\right| > \varepsilon\right) \\ &\quad + P\left(\left|\sum_{j=1}^{r} W_T(j)/(rph_1h_2)^{1/2}\right| > \varepsilon\right). \end{aligned}$$

Then if we choose r, p, and q such that $P\left(\left|\sum_{j=1}^{r}(V_T(j) - W_T(j))/(rph_1h_2)^{1/2}\right| > \varepsilon\right) \to 0$, for example, $r \simeq T^a$, $p \simeq T^{1-a}$, $q \simeq T^c$, $0 < a < 1$, $0 < c < 1$,

$$\sum_{j=1}^{r} (V_T(j) - W_T(j))/(rph_1h_2)^{1/2} \xrightarrow{p} 0.$$

For the part $\left|\sum_{j=1}^{r} W_T(j) / (rph_1h_2)^{1/2}\right|$, we show the asymptotic normality as follows. We need to show the following Lyapunov condition.

$$L = \frac{\sum_{j=1}^{r} E|W_T(j)|^3}{(r\text{var}(W_T(j)))^{3/2}} \xrightarrow{p} 0.$$

First, using the result of Lemma 2, it can be shown

$$\text{var}(W_T(j)) = \text{var}(V_T(j)) \sim ph_1h_2 \sum_{i=1}^{k} \lambda_i^2.$$

On the other hand, A.5 implies $EW_T^4(j) = O\left(p^4h_1^4h_2^4\right)$.Then, $L = O\left(r^{-1/2}ph_1h_2\right)$ which tends to zero with appropriate choices of r and p. Also, $\sum_{j=1}^{r} V_T'(j) / (rph_1h_2)^{1/2}$ can be shown easily to tend to zero in probability. Collecting the results thus far yields (43). With (30), this completes the proof. Additionally, since $\hat{f}_u(x_i) - f_u(x_i) \xrightarrow{p} 0$ for $i = 1, \ldots, k$. $f_u(x_i)\{||K||_2^2\}^2$ can be consistently estimated by $\hat{f}_u(x_i)\{||K||_2^2\}^2$. ■

Lemma 5 *Let I be any compact subset of $\mathbb{R}$. Assume that bandwidth h_1, h_2 are chosen such that $(\ln T) / (Th_1h_2) = o(1)$ as $T \to \infty$. Then,*

$$\sup_{x \in I, u \in [\delta_T, 1-\delta_T]} \left|\hat{f}(u,x) - g_0(u,x)\right| = O_p\left(h_1^2 + h_2^2\right) + O\left[\left(\frac{\ln T}{Th_1h_2}\right)^{1/2}\right].$$

Proof of Lemma 5. See Hansen (2008) with Lemma 2 and Theorem 1 of Kristensen (2009). ■

Appendix C. Option pricing application

Analogous to Aït-Sahalia (1996), we could develop option pricing estimation method in the vicinity of a time point, u. In the neighbourhood of a time point, u, let $\mathcal{A}_u(t,x)$ be the price of the derivative of interest with current time t and maturity date T where x is the underlying asset price. Also, define $\gamma_u(x)$, $c_u(t,x)$ and $b_u(x)$ be the market price of interest rate risk, the cash flow rate from the corresponding derivative security per unit of time, and the payoff of the derivative at maturity T respectively.

$$L\mathcal{A}_u(t,x) = -c_u(t,x)$$

where L is the parabolic differential operator.

Due to Itô-Lemma,

$$
\begin{aligned}
d\mathcal{A}_u(t,x) &= \frac{\partial \mathcal{A}_u}{\partial t}dt + \frac{\partial \mathcal{A}_u}{\partial x}dx + \frac{1}{2}\frac{\partial^2 \mathcal{A}_u}{\partial x^2}(dx)^2 \\
&= \frac{\partial \mathcal{A}_u}{\partial t}dt + \frac{\partial \mathcal{A}_u}{\partial x}(\mu(\cdot)dt + \sigma(\cdot)dW_t) + \frac{1}{2}\frac{\partial^2 \mathcal{A}_u}{\partial x^2}\sigma^2(\cdot)dt \\
&= \left(\frac{\partial \mathcal{A}_u}{\partial t} + \frac{\partial \mathcal{A}_u}{\partial x}(\beta(\cdot)(\alpha(\cdot)-x)) + \frac{1}{2}\frac{\partial^2 \mathcal{A}_u}{\partial x^2}\sigma^2(\cdot)\right)dt + \frac{\partial \mathcal{A}_u}{\partial x}\sigma(\cdot)dW_t
\end{aligned}
$$

On the other hand, due to martingale property, no arbitrage condition yields

$$
\frac{d\mathcal{A}_u}{\mathcal{A}_u} = m(t,\mathcal{A}_u)dt + s(t,\mathcal{A}_u)dW_t,
$$

where $m(t,\mathcal{A}_u) = x - \frac{c(t,x)}{A(t,x)} + \gamma(t,x)s(t,\mathcal{A}_u)$, $s(t,\mathcal{A}_u)\mathcal{A}_u = \frac{\partial \mathcal{A}_u}{\partial x}\sigma(\cdot)$ and $\gamma(\cdot)$ is the market price of interest rate risk. Therefore, the following relationship holds.

$$
\left(\frac{\partial \mathcal{A}_u}{\partial t} + \frac{\partial \mathcal{A}_u}{\partial x}(\beta(\cdot)(\alpha(\cdot)-x)) + \frac{1}{2}\frac{\partial^2 \mathcal{A}_u}{\partial x^2}\sigma^2(\cdot)\right) = \mathcal{A}_u\left\{x - \frac{c(t,x)}{\mathcal{A}_u(t,x)} + \gamma(t,x)\frac{\partial \mathcal{A}_u}{\partial x}\sigma(\cdot)\right\}. \tag{44}
$$

It is worth noting that (44) holds for any asset $\mathcal{A}_u$. [9]

Since the left hand side of the above equation can be written as

$$
\begin{aligned}
L\mathcal{A}_u \equiv & -\partial\mathcal{A}_u/\partial t + \left[\sigma_u^2(x)/2\right]\left(\partial\mathcal{A}_u^2/\partial x^2\right) \\
& + \left[\mu_u(x;\theta) - \gamma_u(x)\sigma_u(x)\right](\partial\mathcal{A}_u/\partial x) - x\mathcal{A}_u.
\end{aligned}
$$

Obviously, $L\mathcal{A}_u(x,T) = b_u(x)$. In mathematics term, this is a Cauchy problem and therefore, there exists a unique solution for the following linear parabolic partial differential equation under certain conditions.

$$
\begin{aligned}
L\mathcal{A}_u(x,T) &= b_u(x) \text{ for all } x \in (0,\infty) \\
L\mathcal{A}_u(t,x) &= -c_u(t,x).
\end{aligned}
$$

Numerical solutions for the above equation could be calculated via the finite-difference method but this method is rather restrictive. Instead, given that Feynman-Kac representation links solutions to the above equation with the conditional moment involving the underlying $\{X_t\}$. In this regard, with appropriate modification and additional assumptions, our estimation of prices of financial derivatives can be valid as long as the corresponding assumptions in Hull and White (1990) and Kristensen (2008) are met. For more theoretical and empirical details, see Hull and White (1990), Stanton (1997) and Kristensen (2008).

[9] In fact, (44) holds under certain conditions. For example, the price of a derivative security of an underlying asset is determined by only the underlying asset itself. For more details, see Vasicek, CIR, and HW.

Appendix D. Tables and Figures

Table 1: Unit root tests

Period, **Type of Tests**	**Global**	**Sep.75-Aug.77**	**Dec.78-Dec.80**	**Jun.82-Jun.84**	**Feb.86-Feb.88**	**Jun.92-May.94**
PP† (-3.4106)‡	-2.9265	-6.9862**	-3.003	-5.1148**	-3.2746⁻	-3.1597⁻
NP (0.0199)	0.0252	0.00857**	0.01420*	0.01236*	0.01391*	0.01389*

*Statistically significant at the 5 percent level.

**Statistically significant at the 1 percent level.

⁻ Statistically significant at the 10 percent level.

† Constant and linear Trend, Bartlett kernel and Newey-West Bandwidth are used.

‡ Values in parentheses denote critical values corresponding to 5 % significance level

Table 2: Option pricing - Nonparametric Underlying Bond Prices

Maturity (Years)	**Annualized Spot Rate (%)**						
	0.02	**0.04**	**0.06**	**0.08**	**0.10**	**0.12**	**0.14**
0.5	0.9829	0.9746	0.9664	0.9585	0.9507	0.9432	0.9358
	0.9807	0.9749	0.9693	0.9635	0.9579	0.9527	0.9471
	0.9712	0.9649	0.9586	0.9525	0.9466	0.9407	0.9346
	0.9799	0.9725	0.9652	0.9583	0.9515	0.9447	0.9380
1	0.9583	0.9448	0.9315	0.9191	0.9075	0.8965	0.8857
	0.9551	0.9478	0.9405	0.9336	0.9269	0.9199	0.9132
	0.9163	0.9075	0.8992	0.8912	0.8834	0.8751	0.8669
	0.9478	0.9367	0.9261	0.9159	0.9059	0.8966	0.8874
3	0.8029	0.7839	0.7662	0.7500	0.7348	0.7195	0.7032
	0.8434	0.8364	0.8295	0.8228	0.8160	0.8094	0.8028
	0.6870	0.6791	0.6712	0.6636	0.6557	0.6480	0.6406
	0.7868	0.7752	0.7639	0.7528	0.7424	0.7320	0.7220
5	0.6590	0.6419	0.6261	0.6122	0.5995	0.5865	0.5743
	0.7473	0.7413	0.7353	0.7292	0.7236	0.7181	0.7126
	0.5107	0.5051	0.4994	0.4937	0.4882	0.4827	0.4773
	0.6387	0.6292	0.6200	0.6113	0.6027	0.5943	0.5859
10	0.3982	0.3887	0.3798	0.3716	0.3634	0.3559	0.3484
	0.5495	0.5449	0.5404	0.5359	0.5316	0.5274	0.5231
	0.2438	0.2412	0.2383	0.2358	0.2333	0.2308	0.2282
	0.4038	0.3977	0.3917	0.3860	0.3808	0.3755	0.3703

1) The face value of the bond of interest is $1.

2) Time t is normalised to 0 so that Time to maturity can be used appropriately. .

3) The four elements in each cell are, from top to bottom, nonparametric call option prices based on estimates of stationary diffusion model, estimates around May 77, estimates around Jul. 80, and estimates around Feb. 93 respectively.

Table 3: Option pricing - Nonparametric Call Option Prices on a 5-year zero coupon Bond

Annualized Spot Rate	**Option Expiration**	**Strike Price 0.98**	**0.99**	**1.00**	**1.01**	**1.02**
0.05	**0.5**	3.0526	2.4396	1.8267	1.2137	0.6008
		3.4755	2.7607	2.0458	1.3310	0.6161
		2.9206	2.4334	1.9462	1.4590	0.9717
		3.1674	2.5589	1.9503	1.3418	0.7333
0.05	**1**	5.2589	4.6665	4.0741	3.4817	2.8893
		5.5186	4.8114	4.1042	3.3970	2.6897
		5.8095	5.3521	4.8947	4.4372	3.9798
		5.3913	4.8015	4.2118	3.6221	3.0323
0.10	**0.5**	4.0509	3.4879	2.9248	2.3618	1.7987
		4.3928	3.7044	3.0159	2.3275	1.6390
		3.7410	3.2713	2.8016	2.3319	1.8622
		4.0513	3.4665	2.8816	2.2968	1.7120
0.10	**1**	6.6773	6.1246	5.5719	5.0192	4.4665
		6.5994	5.9336	5.2679	4.6022	3.9365
		6.8125	6.3665	5.9206	5.4746	5.0287
		6.7942	6.2415	5.6888	5.1361	4.5835
0.15	**0.5**	4.8303	4.3019	3.7735	3.2452	2.7168
		5.2179	4.5484	3.8789	3.2094	2.5399
		4.1633	3.7258	3.2884	2.8509	2.4135
		4.9417	4.3917	3.8417	3.2917	2.7417
0.15	**1**	8.0288	7.5271	7.0255	6.5239	6.0222
		7.7171	7.0754	6.4338	5.7921	5.1504
		7.7594	7.3449	6.9304	6.5158	6.1013
		8.0377	7.5225	7.0073	6.4921	5.9768

1) All of the above values correspond to the prices of call options whose remaining expiration, the spot interest rate at the time of selling and strike prices are denoted in the table. The strike price is expressed as a proportion of the corresponding underlying bond price at the time of selling.

2) The four elements in each cell are, from top to bottom, nonparametric call option prices based on estimates of stationary diffusion model, estimates around May 77, estimates around Jul. 80, and estimates around Feb. 93 respectively.

Table 4: Simulation Results : Point Estimates and Mean Square Error

Observations T		**1000**			**3000**		
Models							
		$\hat{\alpha}(u)$	$\hat{\beta}(u)$	$\hat{\sigma}(u,x)$	$\hat{\alpha}(u)$	$\hat{\beta}(u)$	$\hat{\sigma}(u,x)$
Example 1	$Mean$	0.2486	0.2854	0.0033	0.2493	0.2609	0.0035
	Var	5.5e-04	0.0063	0.00184	0.0002	0.0025	0.00124
$\sigma(u,x) =$	$Bias$	(-0.0013)	(0.07082)	(0.00016)	(-0.0007)	(0.0463)	(0.00010)
$\tau(0.5\cos(0.5\pi u)+1)\sqrt{X_t}$	MSE	(5.5e-04)	(0.01128)	(0.00184)	(0.0002)	(0.0046)	(0.00124)
Example 2	$Mean$	0.2491	0.2580	7.14e-04	0.2495	0.2308	7.33e-04
	Var	1.4e-04	0.0065	4.3e-04	6.0e-05	0.0023	3.2e-05
$\sigma(u,x) =$	$Bias$	(-0.0008)	(0.0433)	(-2.0e-05)	(-0.0004)	(0.0162)	(-2.0e-06)
$\tau 0.5(\cos(0.5\pi u)+(\sin(0.5\pi u)))\sqrt{X_t}$	MSE	(1.4e-04)	(0.0084)	(4.3e-04)	(6.0e-05)	(0.0026)	(3.2e-05)
Example 3	$Mean$	0.2477	0.2611	0.0019	0.2483	0.2321	0.0019
	Var	3.5e-04	0.0064	3.9e-04	1.41e-04	0.00255	0.0002
$\sigma(u,x) =$	$Bias$	(-0.00233)	(0.0464)	(0.000035)	(-0.0016)	(0.00175)	(0.00002)
$\tau(0.25u+1)\sqrt{X_t}$	MSE	(3.6e-04)	(0.00854)	(3.9e-04)	(1.44e-04)	(0.00256)	(0.0002)
Example 4	$Mean$	0.0862	0.2617	0.00073	0.0856	0.2323	0.00051
	Var	0.0001	0.0094	5.0e-06	4.0e-05	0.0023	1.0e-06
$\sigma(u,x) =$	$Bias$	(0.0005)	(0.0471)	(0.0002)	(-0.0001)	(0.0177)	(0.00001)
$\tau\sqrt{X_t}$	MSE	(0.0001)	(0.0094)	(5.0e-06)	(4.0e-05)	(0.0026)	(1.0e-06)

1) 500 repetitions.

2) Parameter values : For Example 1, 2, and 3, $\alpha(u) = -u^2 + u,$, $\beta(u) = 0.2146$, $\tau = 0.0783$ and $u = 0.5, x = 24\%$.

For Example 4, $\alpha = 0.0857$, $\beta = 0.2146$, $\tau = 0.0783$ and $u = 0.5, x = 8\%$.

Table 5: Simulation Results : Integrated Bias (IBias), Variance (IVar) and Mean Square Error (MISE)

Observations T		**1000**				**3000**			
Models									
		$\hat{\alpha}(u)$	$\hat{\beta}(u)$	$\hat{\sigma}(u,\cdot)$	$\hat{\sigma}(\cdot,x)$	$\hat{\alpha}(u)$	$\hat{\beta}(u)$	$\hat{\sigma}(u,\cdot)$	$\hat{\sigma}(\cdot,x)$
Example 1	$IVar$	3.164e-04	0.0053	0.0038	0.0014	1.261e-04	0.0020	0.0012	0.0006
	$IBias$	5.064e-06	0.0019	8.29e-08	6.32e-09	1.14e-06	8.968e-04	4.96e-08	3.257e-09
	$MISE$	(3.215e-04)	(0.0072)	(0.0038)	(0.0014)	(1.272e-04)	(0.0029)	(0.0012)	(0.0006)
Example 2	$IVar$	7.841e-05	0.0037	0.0042	0.0014	3.253e-05	0.0014	0.0011	0.0004
	$IBias$	3.152e-06	0.0022	1.47e-08	3.754e-09	7.719e-07	0.0013	5.49e-09	9.83e-10
	$MISE$	(8.156e-05)	(0.0059)	(0.0042)	(0.0014)	(3.33e-05)	(0.0027)	(0.0011)	(0.0004)
Example 3	$IVar$	2.192e-04	0.0043	0.0043	0.0017	9.454e-05	0.0010	0.0018	0.0007
	$IBias$	3.77e-06	5.88e-04	2.94e-08	5.97e-09	1.16e-06	1.606e-04	6.31e-09	3.19e-09
	$MISE$	(2.23e-04)	(0.0049)	(0.0043)	(0.0017)	(9.57e-05)	(0.0012)	(0.0018)	(0.0007)
Example 4	$IVar$	7.925e-05	0.0055	9.8e-04	0.0007	3.162e-05	0.0020	3.51e-04	0.0003
	$IBias$	3.53e-07	0.0012	1.48e-08	3.34e-09	2.77e-07	(1.91e-04)	7.35e-09	9.48e-10
	$MISE$	(7.96e-05)	(0.0067)	(9.8e-04)	(0.0007)	(3.19e-05)	(0.0022)	(3.51e-04)	(0.0003)

1) 500 repetitions. Parameter values are given in Table 1.

2) $MISE(\hat{\theta}) \equiv \int MSE(\hat{\theta})dx$ and $IBias(\hat{\theta}) \equiv \int Bias^2(\hat{\theta})dx$.

3) $\hat{\sigma}(u,\cdot)$ are obtained by integrating the relevant MSE or Bias over x when u is fixed at 1/2.

4) $\hat{\sigma}(\cdot,x)$ are obtained by integrating the relevant MSE or Bias over u when x is fixed at 24% for Example 1,2,3 and at 8% for Example 4.

Figure 1: Movement on the US Treasury Bill (1955.7 - 2008.12)

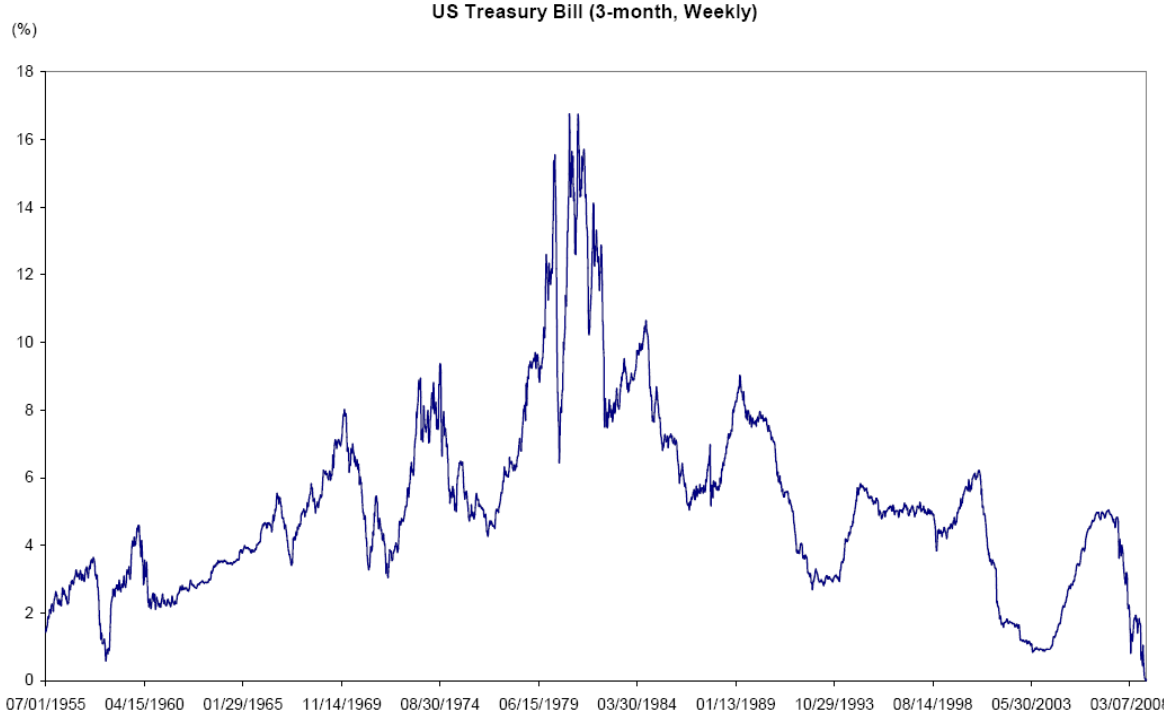


Figure 2: Movement of Eurodollar Deposit Rate (1973.6 - 1995.2)

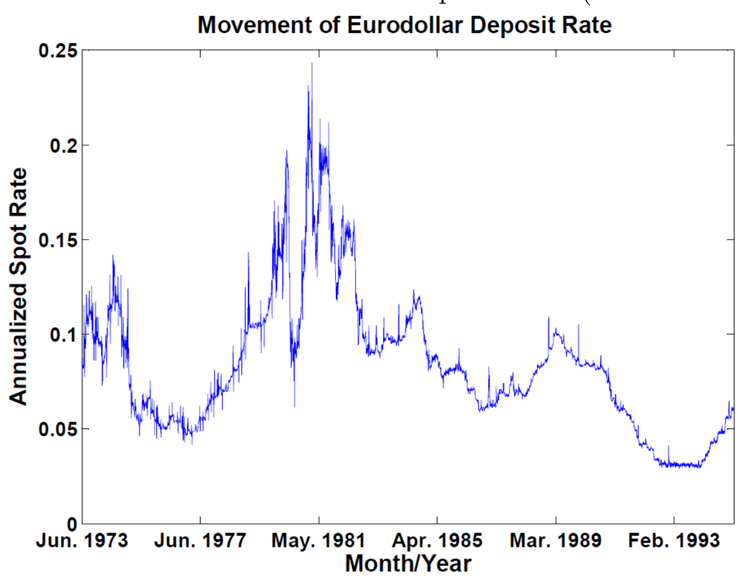
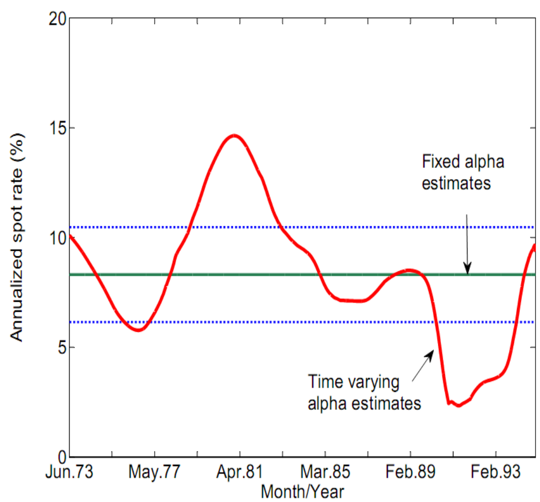
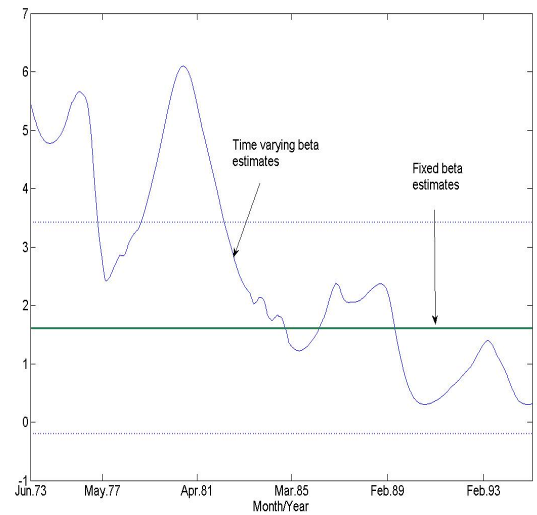


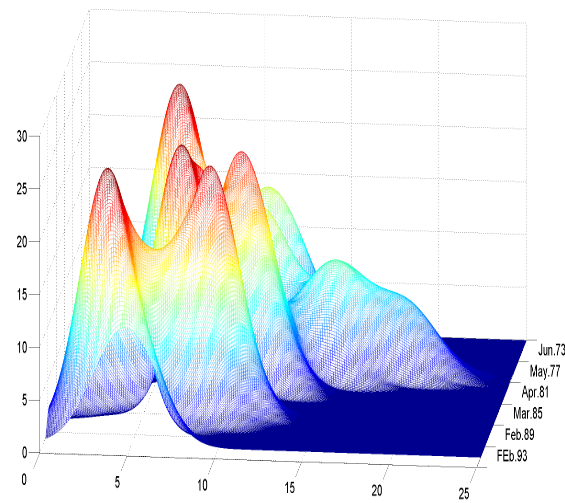
Figure 3: Estimates of $\hat{\mu}$, $\hat{f}$



(a)



(b)



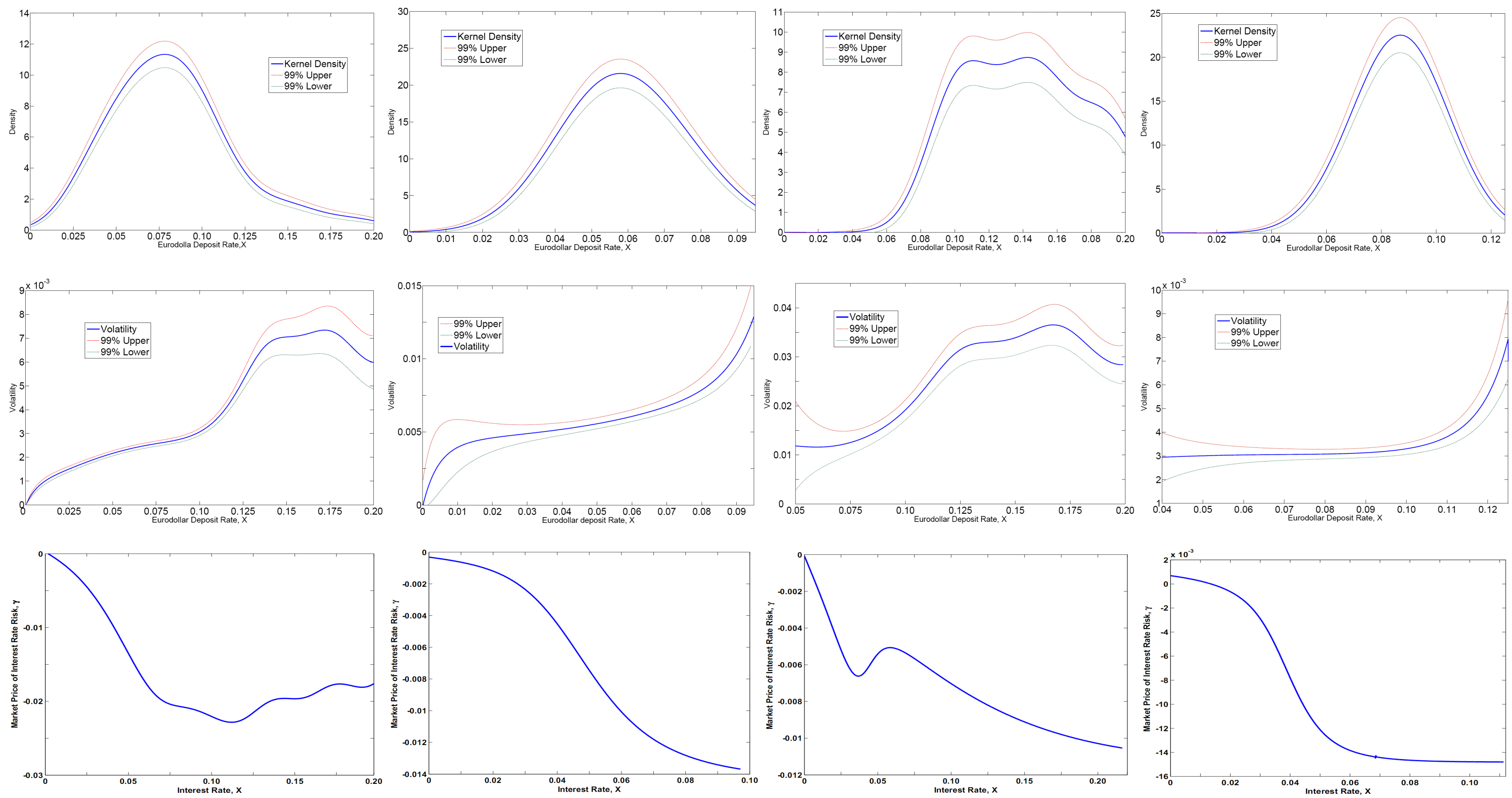
(c)

(a) The time series plots of estimates of time varying α and a fixed α along with 99% upper and lower pointwise confidence interval of the fixed α.

(b) The time series plots of estimates of time varying β and a fixed β along with 99% upper and lower pointwise confidence interval of the fixed β.

(c) Three dimensional surface plots of the estimates of the density function over x and u.

Figure 4: Estimates of $\hat{f}, \hat{\sigma}^2, \hat{\gamma}$



1) The four columns correspond to, from left to right, estimates of Density, Volatility, and Market price of risk (a) stationary diffusion model (Aït -Sahalia, 1996) ; (b) around May 77; (c) around July 80; (d) around February 93.

2) The three elements in each column are, from top to bottom, estimates of density, volatility, and market price of risk, respectively.

3) Dashed (Dotted) lines represent 99% upper (lower) pointwise confidence intervals without the bias correction.