The Spatial Analysis of Time Series

Joon Y. Park[1]
Department of Economics
Rice University and Sungkyunkwan University

Abstract

In this paper, we propose a method of analyzing time series, called the *spatial analysis*. The analysis consists mainly of the statistical inference on the distribution given by the expected local time, which we define to be the *spatial distribution*, of a given time series. The spatial distribution is introduced primarily for the analysis of nonstationary time series whose distributions change over time. However, it is well defined for both stationary and nonstationary time series, and reduces to the time invariant stationary distribution if the underlying time series is indeed stationary. The spatial analysis may therefore be regarded as an extension of the usual inference on the distribution of a stationary time series to accommodate for nonstationary time series. In fact, we show that the concept of the spatial distribution allows us to extend many notions and ideas built upon the presumption of stationarity and make them applicable also for the analysis of nonstationary data. Our approach is nonparametric, and imposes very mild conditions on the underlying time series. In particular, we allow for the observations generated from a wide class of stochastic processes with stationary and mixing increments, or general markov processes including virtually all diffusion models used in practice. For illustration, we provide some empirical applications of our methodology to various topics such as the risk management, distributional dominance and option pricing.

First Draft: December, 2003
This version: August, 2005

JEL classification: C14, C22
Key words and phrases: local time, spatial distribution, semimartingale, markov process, diffusion, risk management, distributional dominance, option pricing

[1]An earlier version of this paper was presented at the 2004 ASSA Meeting, San Diego. I would like to thank Yoosoon Chang and Bandi Federico for many helpful discussions and comments on earlier versions of this paper, and to Minsoo Jeong for excellent research assistance. I am also grateful for their useful comments to the seminar participants at NYU, Caltech/UCLA/USC/UCSD Joint Conference, University of Maryland, LSE, Cass Business School, Leeds University, Montreal Econometrics Group, Harvard/MIT Joint Seminar, Chicago Business School, Boston University, Indiana University and Guanghua School of Management at Peking University.

1. Introduction

In this paper, we develop a new framework to analyze economic and financial time series data, which we call the *spatial analysis* of time series. The spatial analysis is built upon the empirical assessment of and inference on the expected value of the local time of the underlying stochastic process that generates the observed time series. It refers to the staistical analysis of the *spatial distribution*, which we define to be a measure on the real line having the expected local time as its density with respect to the Lebesgue measure, for the stochastic process that yields the given time series observations. As is well known, the local time of a stochastic process measures the sojourn time that it spends in a neighborhood of each spatial point. Therefore, we may easily see that the spatial distribution defined from the expected local time would represent the expected frequency at which the underlying stochastic process visits each spatial point.

The spatial analysis is developed primarily for the time series that are nonstationary, i.e., the time series that do not have time invariant stationary distributions. Many problems in economics and finance are intrinsically of dynamic and time-varying nature. Neverthless, they have been routinely analyzed within a static and time-invariant framework of stationarity. The stationarity, though it is a very convenient and useful concept from the theoretical point of view, is highly unrealistic and unlikely to hold in many cases of the time series studies on economic and financial markets. Unlike the time invariant distribution that exists only under stationarity, the spatial distribution is well defined for general nonstationary, as well as stationary, time series. For both stationary and nonstationary time series models, we may therefore analyze the spatial distribution to make inferences on their distributional characteristics. This is the main motivation of the spatial analysis.

The spatial distribution reduces to the time invariant stationary distribution, if the underlying time series is indeed stationary. The spatial analysis may therefore be viewed as a generalization of the conventional inference on the time invariant distribution of stationary time series. In fact, the concept of spatial distribution allows us to extend various notions and ideas that have been developed under the presumption of stationarity, and make them applicable for nonstationary time series as well. For a general nonstationary time series, the spatial distriblution may be interpreted as the aggregate of its time-varying distribution over a period of time. Moreover, the sum of the expected utilities generated by a stochastic process is determined soley by its spatial distribution, and therefore, it may well be conceived that the spatial distribution plays the central role in many economic and financial problems involving dynamic decision making based on utility maximization.

The time series data are routinely plotted on the xy-plane with "x" and "y" representing respectively the time and spatial axes. The usual readings of data along the time axis is truly meaningful only under the assumption of stationarity, which allows us to interpret those readings as repeated observations from the common underlying distribution. Clearly, this interpretation is not possible for nonstationary data whose distributions are changing over time. The readings along the spatial axis can, however, be meaningful for nonstationary data, as well as for stationary data. In particular, they can be very useful for the analysis of the time series which take values repeatedly over a certain range, like many economic and financial time series. Roughly, such data can be read as repeated observations along the spatial axis. This is what makes our spatial analysis applicable for a wide class of stationary and nonstationary economic and financial data.

The spatial analysis can be very useful for a wide variety of economic and financial studies. For the illustrative purpose, we explore a few of such possibilities in the paper. First, we obtain a new measure of risk, called the aggregate value-at-risk (VaR), under a more realistic assumption that the value of the underlying asset is driven by a stochastic process having distributions changing over time. Second, we introduce the notion of spatial dominance, which generalizes the stochastic dominance. The former compares the expected sums of utilities over time in a general dynamic and nonstationary setting, while the latter only allows us to look at the expected utilities either at a given fixed time in a completely static setting or under the assumption of strict stationarity. The spatial dominance can be used to rank investment strategies and socio-economic programs that need to be evaluated over a certain period of time. Third, we note that our spatial analysis can also be used for the option pricing. As shown by Carr and Jarrow (1990), the arbitrage-free price of a European option is given by the expected local time of the underlying asset price evaluated at exercise price.

We provide in the paper the statistical tools and methodologies that are useful for the spatial analysis. Their asymptotics are also fully developed. In particular, we establish the consistency of the proposed estimators and obtain their limit distributions. Moreover, we develop various statistics that can be used to test for many interesting hypotheses on the spatial distributions for the multiple time series data. The critical values are in general dependent upon the distributions of the underlying stochastic processes, but they may be readily calculated via simulation, bootstrap or sub-sampling method. Our framework is very general, requiring only very weak conditions on the underlying stochastic processes. More precisely, we develop two sets of methodologies, one for the stochastic processes with stationary and mixing increments and the other for the general markov processes. The results in the paper should therefore be applicable for a very wide class of stochastic processes including virtually all models that are used in practical applications.

For the actual spatial analysis, we explicitly look at several statistical procedures. First, we provide a method to obtain the point forecast of spatial distribution with confidence band. This can be used to predict the aggregated distribution of a time series over some future period of time. Second, we investigate the specification test in the spatial domain. Here we intend to test whether the time series of interest has a certain spatial distribution. Third, the test of equality in spatial distributions is also considered. It tests whether or not two time series have a common spatial distribution, or equivalently, they are indistinguishable in the spatial domain. Lastly, the statistical test of spatial dominance is proposed. Analogously as for stochastic dominance, a time series is said to spatially dominate the other if it has the spatial distribution that dominates that of the other. We only consider the test of the first-order dominance, but it easily extends to other types of dominances.

The rest of the paper is organized as follows. We present some motivations and preliminaries in Section 2. The notion of local time is introduced and extended to define several variants of local time, which will be used for our subsequent spatial analysis. Section 3 provides some immediate practical applications of spatial analysis on various topics including risk analysis, comparing expected utilites and option pricing. Section 4 lays out some fundamentals of our asymptotic analysis. Our asymptotic framework is introduced, and the basic asymptotics for the estimators of local time and its variants are developed there. Section 5 provides the estimators of the spatial distributions, and derive their asymptotic properties. The methodologies and theories are given in sequel for the processes with stationary increments and general markov processes. In Section 6, we study various inferences that we may utilize for the spatial analysis of time series. It includes forecast of spatial distribution, specification test in spatial domain, test of equality in spatial distribution and test of spatial dominance. The concluding remarks follow in Section 7. All the mathematical proofs of the theorems in the paper are given in Appendix.

2. Motivations and Preliminaries

We let $X = (X_t)$ be a stochastic process. If we denote by μ the Lebesgue measure on $\mathbb{R}$ and let the sojourn time ν of X in any Borel set $A \subset \mathbb{R}$ up to time T be given by $\nu(T, A) = \mu\{t \in [0, T] | X_t \in A\}$, then the *local time* of X is formally defined as the Radon-Nykodim derivative of $\nu(T, \cdot)$ with respect to μ, i.e.,

$$\ell(T, x) = \frac{d\nu}{d\mu}(T, x), \tag{1}$$

where we assume that $\nu(T, \cdot)$ is absolutely continuous with respect to μ.[2] Consequently, for any Borel set $A \subset \mathbb{R}$, the integral of $\ell(T, \cdot)$ over A yields the sojourn time of X in A up to time T. The local time $\ell(T, x)$ therefore represents the frequency at which the process X visits the spatial point x up to the time T. As is obvious from this definition, the local time ℓ itself is a stochastic process defined on the underlying stochastic process X. It has two parameters, T and x, which we will refer respectively to the time and spatial parameters. The reader is referred to, e.g., Bosq (1998) and Revuz and Yor (1994) for more discussions on the local time.

From the definition of local time in (1), it follows immediately that

$$\int_0^T u(X_t)\, dt = \int_{-\infty}^{\infty} u(x)\ell(T, x)\, dx \tag{2}$$

for any nonnegative Borel-measurable function u on $\mathbb{R}$. This is well known and often referred to as the occupation times formula. If the local time $\ell(T, \cdot)$ of X is continuous, then we may easily deduce from the occupation times formula that

$$\ell(T, x) = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^T 1\{|X_t - x| < \varepsilon\}\, dt, \tag{3}$$

which makes it clear that $\ell(T, \cdot)$ can be regarded as a 'density'. Clearly, we may also apply the occupation times formula with the choice of $u(y) = 1\{y \leq x\}$ and obtain the corresponding 'distribution function'

$$L(T, x) = \int_{-\infty}^{x} \ell(T, y)\, dy = \int_0^T 1\{X_t \leq x\}\, dt, \tag{4}$$

[2]For a semimartingale X, the sojourn time is usually measured by the quadratic variation. Our definition here, however, is more convenient for our subsequent analysis.

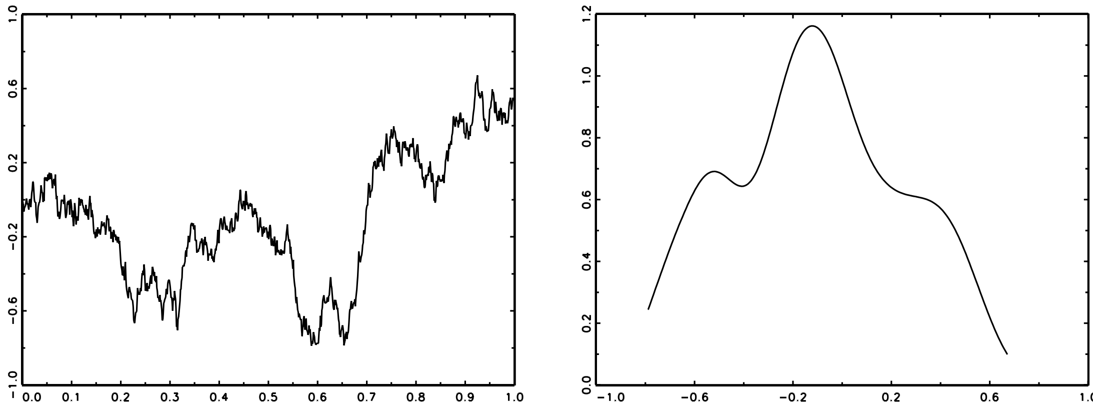

Figure 1: Simulated Sample Path and Estimated Local Time of Brownian Motion

which will be called the *integrated local time* in the paper.

The local time is well defined for a broad class of stochastic processes including all continuous semimartingales. Moreover, most of the stochastic processes that are used in practical applications have a version of the local time that is continuous with respect to the spatial parameter. For any continuous local martingale, we may indeed choose a version of ℓ so that $\ell(T,\cdot)$ is Hölder continuous of order $1/2-\delta$ for any $\delta>0$ [see, e.g., Theorem 1.2, Corollaries 1.8 and 1.9 in Chapter VI of Revuz and Yor (1994)]. The Hölder continuity may also be established for the local times of more general continuous semimartingales under some mild extra regularity conditions [see, e.g., Exercise 1.32 in Chaper VI of Revuz and Yor (1994) for details].[3] The existence and property of the local time will not be further discussed in the paper. Instead, we will just assume that the underlying stochastic process is a semimartingale, for which the local time is well defined and continuous with respect to the spatial parameter so that in particular our representation in (3) is valid. See Figure 1 for a realization of the standard Brownian motion and the estimated local time.

As mentioned earlier, the local time ℓ itself is a stochastic process and random. We may therefore take the expectation and define

$$\lambda(T,x)=\mathbb{E}\,\ell(T,x)=\lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T \mathbb{P}\{|X_t-x|<\varepsilon\}\,dt \tag{5}$$

and

$$\Lambda(T,x)=\mathbb{E}\,L(T,x)=\int_0^T \mathbb{P}\{X_t\le x\}\,dt. \tag{6}$$

Throughout the paper, we call λ and Λ respectively the *spatial density* and the *spatial distribution function*.[4] Naturally, we may define the *spatial distribution* to be the distribu-

[3] In the literature, the continuity properties are given for the local time defined in terms of the quadratic variation $[X]$ of the underlying stochastic process X. They can, however, be readily translated into those for the version of local time defined in our paper if we specify $[X]$ more explicitly.

[4] It is also worth noting that $(\partial/\partial t)\Lambda(t,x)=\mathbb{P}\{X_t\le x\}$, which can be deduced directly from (6) and the fundamental theorem of calculus.

tion given by the spatial density or the spatial distribution function.[5] Of course, we have $\Lambda(T,x) = \int_{-\infty}^{x} \lambda(T,y)\,dy$ by Fubini's theorem. Our subsequent discussion will be mainly focused on the *spatial analysis*, i.e., the statistical analysis of the spatial distribution, of the stochastic process generating a given time series.

For the spatial density, we may well expect that the result corresponding to the occupation times formula in (2) holds.

Lemma 2.1 *We have*

$$\mathbb{E}\int_0^T u(X_t)\,dt = \int_{-\infty}^{\infty} u(x)\lambda(T,x)\,dx$$

for any nonnegative Borel-measurable function u on $\mathbb{R}$.

Lemma 2.1 shows in particular that, for any given utility function, the sum of expected future utilities generated by a stochastic process over a period of time is determined by, and only by, its spatial distribution. It is therefore not difficult to see that the spatial distribution plays the central role in analyzing many dynamic maximization problems in economics, finance and other related fields involving the expected future utilities. This will be demonstrated more clearly in the next section.

If the underlying process X is stationary and has the time invariant continuous density π and distribution function Π, then it follows immediately from (5) and (6) that

$$\pi(x) = \frac{\lambda(T,x)}{T} \quad \text{and} \quad \Pi(x) = \frac{\Lambda(T,x)}{T},$$

since, for each $x \in \mathbb{R}$, $\lim_{\varepsilon\to 0}(1/2\varepsilon)\mathbb{P}\{|X_t - x| < \varepsilon\} = \pi(x)$ and $\mathbb{P}\{X_t \leq x\} = \Pi(x)$ are time invariant and identical for all $t \in [0,T]$. The analysis of spatial distribution would thus reduce in this case to that of time invariant stationary distribution of the underlying process. Therefore, our spatial analysis can be viewed as a natural extension of the statistical analysis on the time invariant distribution for the stationary process to that on a more general spatial distribution for the possibly nonstationary process. For the nonstationary process, we may simply regard $\lambda(T,\cdot)$ and $\Lambda(T,\cdot)$ respectively as the density and distribution function for the distribution of the values of X, which is nonstationary and time varying, aggregated over time $[0,T]$.

For some special stochastic processes, the distribution of the local time is known and we may therefore analytically obtain the spatial density and distribution function. The leading example is Brownian motion. If we denote respectively by φ and Φ the density and distribution function of standard normal distribution, then the spatial density and distribution function of the standard Brownian motion are given by

$$\lambda(T,x) = 2\sqrt{T}\varphi\left(\frac{x}{\sqrt{T}}\right) - 2|x|\Phi\left(-\frac{|x|}{\sqrt{T}}\right)$$

[5] Strictly speaking, the spatial distribution is not a probability measure unless $T = 1$. We may, however, consider the distribution given by $\lambda(T,x)/T$ and $\Lambda(T,x)/T$, whenever it is more convenient to define the spatial distribution as a probability measure.

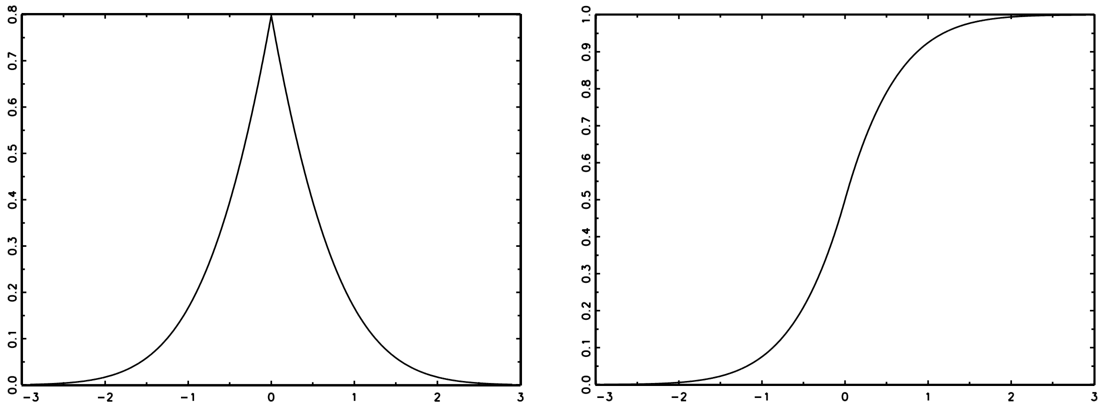

Figure 2: Spatial Density and Distribution Function of Brownian Motion

and

$$\Lambda(T,x) = T\Phi\left(\frac{x}{\sqrt{T}}\right) + x\sqrt{T}\varphi\left(\frac{x}{\sqrt{T}}\right) - (\operatorname{sgn} x)x^2\Phi\left(-\frac{|x|}{\sqrt{T}}\right)$$

for $x \in \mathbb{R}$. This can be obtained readily from Borodin (1989, p5) after some straightforward derivation. For the standard Brownian motion, $\lambda(T,\cdot)$ is symmetric around the origin, and consequently, we have in particular the relationships $\lambda(T,-x) = \lambda(T,x)$ and $\Lambda(T,-x) = T - \Lambda(T,x)$ for all $x \in \mathbb{R}$. The spatial density and distribution function of the standard Brownian motion for $T = 1$ are shown in Figure 2.

To deal with more general dynamic decision making problems, we may consider the *discounted local time*, or simply *d-local time*, ℓ^r of X, which is defined as

$$\ell^r(T,x) = \int_0^T e^{-rt}\ell(dt,x) \tag{7}$$

for some discount rate $r > 0$. The local time ℓ is monotone increasing with respect to the time parameter, and therefore, the discounted local time is well defined. Obviously, the *integrated d-local time* of X given by $L^r(T,x) = \int_0^T \ell^r(T,y)\,dy = \int_0^T e^{-rt}1\{X_t \leq x\}\,dt$ may be defined accordingly. For our spatial analysis, we also introduce the *discounted spatial density*, or *d-spatial density* for short, of X given by

$$\lambda^r(T,x) = \mathbb{E}\ell^r(T,x) = \int_0^T e^{-rt}\lambda(dt,x) \tag{8}$$

corresponding to our definition of the d-local time in (7). Moreover, we let the *d-spatial distribution function* of X be defined as $\Lambda^r(T,x) = \int_0^x \lambda^r(T,y)\,dy = \int_0^T e^{-rt}\mathbb{P}\{X_t \leq x\}\,dt$. In what follows, we refer to the distribution given by the d-spatial density and distribution function as the *d-spatial distribution*.

Corollary 2.2 *We have*

$$\mathbb{E}\int_0^T e^{-rt}u(X_t)\,dt = \int_{-\infty}^{\infty} u(x)\lambda^r(T,x)\,dx$$

for any nonnegative Borel-measurable function u on $\mathbb{R}$.

Corollary 2.2 allows us to consider the sum of expected future utilities discounted by the subjective rate r of time preference. It extends the result in Lemma 2.1, and shows that the discounted expected future utilities generated by a stochastic process is completely determined by the d-spatial distribution.

In this section, we assume that the stochastic process X starts at the origin, i.e., $t = 0$, and consider its spatial distribution over the time interval $[0, T]$. This convention will be made throughout the paper, unless stated explicitly otherwise, for all the stochastic processes that are analyzed. It should also be emphasized that the probability $\mathbb{P}$ and expectation $\mathbb{E}$ here and elsewhere in the paper are to be understood as the conditional probability and expectation given the values of the underlying stochastic processes at the origin. All our statistical methods including forecasts and hypotheses tests are developed primarily for nonstationary processes whose distributions are in particular dependent upon their starting values. It is therefore well predicted that the theories for all our methodologies rely on the initial values of the underlying stochastic processes in a very essential manner. However, for the notational simplicity, we suppress also in the rest of the paper the dependencies of $\mathbb{P}$ and $\mathbb{E}$ on the initial values of the underlying stochastic processes.

3. Illustrative Examples

Before introducing the statistical methods and theories that are needed to implement our new notion and methodology, we discuss some important practical applications to which they can be immediately applied. Presented below are some prototypical examples covering several topics including the risk analysis, distributional dominance and option pricing. The examples are selected for the purpose of illustration. Clearly, many other related problems can be analyzed similarly as the examples given here.

3.1 Risk Analysis

It is customary to measure the risk in a portfolio of financial assets using the concept of the value-at-risk (VaR). As is well known, the value-at-risk is the loss that will not be exceeded at the chosen confidence level. For example, with a confidence level $(1 - \alpha)$, the VaR corresponds to the α-percentile point on the distribution of gains and losses. Let the changes in the portfolio value follow a stationary stochastic process X that has a time invariant distribution function Π. The VaR associated with the confidence level $(1 - \alpha)$ is then given by x_α such that

$$\Pi(x_\alpha) = \mathbb{P}\{X_t \leq x_\alpha\} = \alpha,$$

which is assumed to be the same for all t. The stationarity in this strict form of the underlying process X, however, is highly unlikely to hold in practice. It is widely believed by both researchers and practioners that the underlying process is nonstationary, and in particular, has variability increasing with time, in many cases.

The spatial analysis naturally extends the concept of the VaR for a nonstationary stochastic process that has distributions changing over time. For the measurement of the risk in a portfolio whose value, net of the present value, is driven by a general nonstationary stochastic process, we may use the spatial distribution of the underlying process. Assume without loss of generality that $T = 1$ in this case. Then the risk in holding the portfolio over time $[0, 1]$ may indeed be measured with a confidence level $(1 - \alpha)$ by the *aggregate VaR*, which is given by x_α such that

$$\Lambda(1, x_\alpha) = \int_0^1 \mathbb{P}\{X_t \leq x_\alpha\}\, dt = \alpha,$$

where Λ is the spatial distribution function (6) introduced in the previous section. Now, α represents the time aggregate of the probabilities that we lose more than x_α over the period $[0, 1]$ of time. Naturally, the aggregate VaR reduces to the conventional VaR, if the changes in portfolio value follow a stationary stochastic process.

3.2 Distributional Dominance

For two stationary processes X and Y with the time invariant densities π^X and π^Y and the distribution functions Π^X and Π^Y, we say that X stochastically dominates Y if and only if

$$\Pi^X(x) \leq \Pi^Y(x) \tag{9}$$

for all $x \in \mathbb{R}$. It is well known that the condition in (9) holds if and only if

$$\mathbb{E}\, u(X_t) \geq \mathbb{E}\, u(Y_t) \tag{10}$$

or, equivalently,

$$\int_{-\infty}^{\infty} u(x)\pi^X(x)\, dx \geq \int_{-\infty}^{\infty} u(x)\pi^Y(x)\, dx \tag{11}$$

for every monotone nondecreasing utility function u. Therefore, if X stochastically dominates Y, then X yields at least the same level of expected utility than Y for any monotone nondecreasing utility function. The concept of stochastic dominance is known to be very useful in ordering investment strategies and welfare outcomes such as income distributions and poverty level, and in various socio-economic program evaluation exercises.

Obviously, the notion of stochastic dominance is not much meaningful for nonstationary processes whose distributions change over time. In this case, we need to consider the inequality for the expected sum of instantaneous utilities

$$\mathbb{E} \int_0^T u(X_t)\, dt \geq \mathbb{E} \int_0^T u(Y_t)\, dt \tag{12}$$

or, equivalently,

$$\int_{-\infty}^{\infty} u(x)\lambda^X(T, x)\, dx \geq \int_{-\infty}^{\infty} u(x)\lambda^Y(T, x)\, dx \tag{13}$$

in place of (10) or (11) to claim that X provides at least the same level of expected utility than Y over a certain period of time $[0, T]$. However, we may show that (12) or (13) holds for any monotone nondecreasing utility function u, if and only

$$\Lambda^X(T, x) \leq \Lambda^Y(T, x) \tag{14}$$

for all $x \in \mathbb{R}$. If (14) holds, we say that X *spatially dominates* Y. It is easy to see that the concept of spatial dominance reduces to that of stochastic dominance if the underlying processes are stationary.

We may also consider the spatial dominance with time discount using the d-spatial distribution introduced earlier. Let $\lambda^{r,X}$ and $\lambda^{r,Y}$ denote respectively the d-spatial densities of X and Y defined in (8), and let $\Lambda^{r,X}$ and $\Lambda^{r,Y}$ be their distribution functions. Then we can show that

$$\mathbb{E}\int_0^T e^{-rt}u(X_t)\,dt \geq \mathbb{E}\int_0^T e^{-rt}u(Y_t)\,dt$$

or, equivalently,

$$\int_{-\infty}^{\infty} u(x)\lambda^{r,X}(T,x)\,dx \geq \int_{-\infty}^{\infty} u(x)\lambda^{r,Y}(T,x)\,dx$$

holds for any monotone nondecreasing utility function u, if and only if

$$\Lambda^{r,X}(T,x) \leq \Lambda^{r,Y}(T,x) \tag{15}$$

is satisfied. If (15) holds, we say that X spatially dominates Y with the rate r of time preference.

3.3 Option Pricing

The spatial analysis can also be used in pricing options.[6] Let X be a stochastic process driving the price of a financial asset, over which a European call option is written with strike price x and maturity T. Also, assume that the quadratic variation process $[X]$ of X has the time derivative, and it is given by

$$d[X]_t = \sigma^2(X_t)\,dt.$$

This assumption is satisfied for a wide class of stochastic processes, including all the diffusion processes. If X is indeed a diffusion, then σ becomes its diffusion function.

Under this setting, the arbitrage-free pricing theory suggests that the price of the option at time $t = 0$ be given by

$$\max\left(0, X_0 - e^{-rT}x\right) + \frac{\sigma^2(x)}{2}e^{-rT}\lambda(T,x),$$

where λ is the spatial density of X in (5) obtained under the probability measure known as the equivalent martingale measure. This was shown earlier by Carr and Jarrow (1990), and follows immediately from the so-called Ito-Tanaka formula [see, e.g., Theorem 1.5 in Chapter 6 of Revuz and Yor (1994) for the details]. Our methodologies that will be developed subsequently for the estimation of and testing for the spatial distribution can therefore be used for pricing options.

[6]The example here was provided and discussed by Bandi Federico at the 2004 ASSA Meeting, San Diego, where an earlier version of this paper was presented.

4. Basic Asymptotic Analysis

Being the expected value of the distribution given by the local time, the spatial distribution may naturally be estimated using the average of the repeated estimates for the distribution given by the local time. The estimation of the local time and its variants is therefore an essential ingradient of our spatial analysis. In this section, we show how we may estimate the local time and its variants, and develop their asymptotic theories. Our results in this section would provide the basic methodologies and theories for the statistical analyses of the spatial distributions, which we will explore in subsequent sections. We first present our asymptotic framework, and then develop the asymptotics for the local time and its variants.

4.1 Asymptotic Framework

For all our subsequent analyses, we assume that the underlying stochastic process X has a.s. continuous sample path with the modulus of continuity

$$\omega(\Delta) = \max_{|t-s|\leq\Delta} |X_t - X_s|$$

for all $s, t \geq 0$. This assumption will also be made for all the stochastic processes that we consider in the paper. Some important remarks on this assumption are now in order.

Remarks (a) The modulus of continuity for Brownian motion was found by Levy, which is given by

$$\omega(\Delta) = \sqrt{\Delta \log(1/\Delta)}$$

[see Karatzas and Shreve (1991, Theorem 9.25, p114)]. This also applies to general diffusions if the drift function is differentiable with locally bounded derivative, and the diffusion function is locally bounded. To see this, we write

$$X_t - X_s = \int_s^t \mu(X_u)\,du + \int_s^t \sigma(X_u)\,dW_u$$

where

$$|\mu(X_u) - \mu(X_s)| \leq |X_t - X_s| \max_{t\leq T} |\mu'(X_t)|$$

and

$$\int_s^t \sigma^2(X_u)\,du \leq |t-s| \max_{t\leq T} \sigma^2(X_t)$$

We thus have the given modulus of continuity from the DDS representation of the diffusion part. This is well known.

(b) If there exists $\alpha, \beta > 0$ such that

$$\mathbb{E}|X_t - X_s|^\alpha \leq c|t-s|^{1+\beta}$$

for all $s, t \leq T$ with some constant $c > 0$, then by Kolmogorov's criterion [see Revuz and Yor (1994, Theorems 1.8 and 2.1, pp18, 25)], X has a modification whose sample path is a.s. Hölder continuous of order $\delta \in [0, \beta/\alpha)$. We may therefore have $\omega(\Delta) = \Delta^\delta$ for any $\delta \in [0, \beta/\alpha)$ in this case.

All our subsequent asymptotics require that $\Delta \to 0$, and that $\omega(\Delta) \to 0$ as $\Delta \to 0$. Throughout the paper, we suppose that we have discrete observations $(X_{i\Delta})$ from a continuous stochastic process X, where Δ denotes the observation interval. For the time interval $[0, T]$, the number of observations is thus given by $n = T/\Delta$. All the asymptotics derived in the paper assume that $n \to \infty$ via $\Delta \to 0$ for a fixed T. Our theories are thus obtained by the 'infill' asymptotics, in contrast to the conventional 'long-span' asymptotics relying on $T \to \infty$. Obviously, the infill asymptotics are more appropriate for the spatial analysis, which intends to statistically analyze the spatial distribution of a time series over a fixed time interval. Besides, the infill asymptotics are crucial to deal with the nonstationarity of the underlying process at the level of generality that we entertain in the paper. Quite clearly, the long-span asymptotics alone cannot generate enough information to make inference on general nonstationary processes. Needless to say, our limit distribution theories are more appropriate to analyze the time series data that are sampled at high frequencies.

4.2 Asymptotics for Estimators of Local Time and Its Variants

Given observations $(X_{i\Delta}), i = 1, \ldots, n$, from $X = (X_t)$, we may consistently estimate the local time of X using a renormalized kernel density estimator. We define the kernel estimator for the local time ℓ of the underlying stochastic process X by

$$\hat{\ell}(T, x) = \frac{\Delta}{h} \sum_{i=1}^{n} K\left(\frac{X_{i\Delta} - x}{h}\right), \tag{16}$$

where K is the kernel function and h is the bandwidth parameter. The kernel local time estimator is nothing but a renormalized version of the standard kernel density estimator. Note that we have $\hat{\pi}_n(x) = \hat{\ell}(T, x)/T$, where $\hat{\pi}_n$ is the usual kernel density estimator given by $\hat{\pi}_n(x) = \hat{\ell}(T, x)/T$ for each $x \in \mathbb{R}$. In our context, $\hat{\pi}_n(x)$ would thus provide an estimate of $\ell(T, x)/T$. If the underlying stochastic process X is stationary and has the time invariant density π, then we may indeed show that $\ell(T, x)/T$ converges a.s. as $T \to \infty$ to $\pi(x)$ for each $x \in \mathbb{R}$ under general regularity conditions. The reader is referred to Bosq (1998, Chapter 6) for more discussions on the estimation of local times for stationary processes. Here we do not assume stationarity. It is simply allowed as a very special and rather trivial case, and we will primarily focus on the local time estimators for nonstationary processes.

For the kernel function K, we assume throughout the paper that it is nonnegative and satisfies the usual conditions for the second-order kernels. Our local time asymptotics require some additional conditions, as we assume in

Assumption 4.1 Let K be (a) infinitely differentiable with bounded and absolutely integrable derivatives, and (b) $\int_{-\infty}^{\infty} sK(s)\,ds = \int_{-\infty}^{\infty} sK^2(s)\,ds = 0$ and $\int_{-\infty}^{\infty} s^4 |K(s)|\,ds < \infty$.

Assumption 4.1 holds for the normal kernel. The conditions in Assumption 4.1 are imposed for convenience to simplify the proofs for our subsequent theoretical results. They can be relaxed if we require more stringent conditions on $\omega(\Delta)$ in relation to h. This can be seen in the proof of Theorem 4.1. In general, we may allow for less smooth kernels if $\omega(\Delta) \to 0$ faster relative to h.

The following theorem establishes the strong pointwise and L^1-consistency of the kernel local time estimator that we introduced in (16) above.

Theorem 4.1 *Suppose that* Assumption 4.1 *holds. Let h be chosen such that $h \to 0$ and $\omega(\Delta)/h^{1+\delta} \to 0$ as $\Delta \to 0$ for some $\delta > 0$. Then we have $\hat{\ell}(T,x) \to_{a.s.} \ell(T,x)$ as $\Delta \to 0$ for each $x \in \mathbb{R}$. Moreover, it follows that*

$$\int_{-\infty}^{\infty} \left| \hat{\ell}(T,x) - \ell(T,x) \right| dx \to_{a.s.} 0$$

as $\Delta \to 0$.

Several authors previously investigated the kernel local time estimators and established their consistency for diffusion models. However, their results are restricted to pointwise consistencies, i.e., the convergence of $\hat{\ell}(T,x)$ to $\ell(T,x)$ for each $x \in \mathbb{R}$, and to diffusion models. Florence-Zmirou (1993, Proposition 2, p796) first established the pointwise L^2-consistency for the discontinuous indicator-based kernel local time estimator for general diffusions. Bandi and Phillips (2003, Theorem 1 and Corollary 1, p250) later proved the pointwise strong consistency of the kernel local time estimators relying on smooth kernel functions for general diffusion models. See also Phillips (2003) and Bandi (2002) for some related discussions about the kernel estimation of local time. Here we extend the existing results in two directions: our results establish global L^1-consistency and are applicable for more general semimartingales.

The integrated local time can readily be estimated by integrating the estimated local time. We will, however, look at the more straightforward sample analogue estimator that is given by

$$\hat{L}(T,x) = \Delta \sum_{i=1}^{n} 1\{X_{i\Delta} \leq x\} \tag{17}$$

for the integrated local time. The estimator in (17) may be more convenient to use in practice and is also somewhat easier to analyze, since it does not involve the smoothing parameter. This is why we look at the estimator in the paper. Neverthless, all our subsequent theories for the integrated local time are also applicable for any smoothed estimator obtained from the kernel local time estimator. The next theorem provides the strong uniform consistency of the sample analogue estimator for the integrated local time.

Theorem 4.2 *Let $\omega(\Delta) \to 0$ as $\Delta \to 0$. Then we have*

$$\sup_{x \in \mathbb{R}} \left| \hat{L}(T,x) - L(T,x) \right| \to_{a.s.} 0$$

as $\Delta \to 0$.

The d-local time and integrated d-local time can be estimated similarly. For the d-local time, we consider

$$\hat{\ell}^r(T,x) = \frac{\Delta}{h}\sum_{i=1}^{n} e^{-ri\Delta} K\left(\frac{X_{i\Delta}-x}{h}\right) \tag{18}$$

accordingly as the kernel local time estimator defined in (16). The estimator given in (18) will be referred to as the kernel estimator of d-local time. For the integrated d-local time, we similarly look at

$$\hat{L}^r(T,x) = \Delta\sum_{i=1}^{n} e^{-ri\Delta} 1\{X_{i\Delta} \le x\} \tag{19}$$

in parallel with (17) introduced above. The estimator given in (19) will subsequently be called the sample analogue estimator of integrated d-local time.

We may easily derive the consistency results for the kernel estimator of d-local time and the sample analogue estimator of integrated d-local time, correspondingly with those in Theorems 4.1 and 4.2.

Corollary 4.3 *Suppose that the conditions in* Theorem 4.1 *hold. We have* $\hat{\ell}^r(T,x) \to_{a.s.} \ell^r(T,x)$ *for any* $r > 0$ *as* $\Delta \to 0$ *for each* $x \in \mathbb{R}$. *Moreover, it follows that*

$$\int_{-\infty}^{\infty} \left|\hat{\ell}^r(T,x) - \ell^r(T,x)\right| dx \to_{a.s.} 0$$

for any $r > 0$ *as* $\Delta \to 0$.

Corollary 4.4 *Suppose that the condition in* Theorem 4.2 *hold. We have*

$$\sup_{x\in\mathbb{R}} \left|\hat{L}^r(T,x) - L^r(T,x)\right| \to_{a.s.} 0$$

for any $r > 0$ *as* $\Delta \to 0$.

Therefore, the consistencies that we established earlier for the kernel estimator of local time and the sample analogue estimator of integrated local time continue to hold respectively for the kernel estimator of d-local time and the sample analogue estimator of integrated d-local time, under the same set of assumptions.

5. Asymptotic Theories of Spatial Estimators

In this section, we provide the estimators for the spatial density and distribution function, and develop their asymptotics. In particular, we establish their consistency and derive their imiting distributions. We consider two classes of models: processes with stationary increments and markov processes. These two classes include virtually all models that are used for the empirical research in economics and finance. They are, however, not mutually exclusive and have a large set of models in common. Indeed, many models that are commonly employed in practical applications belong to both classes.

5.1 Processes with Stationary Increments

Let τ_k, $k = 0, \ldots, N-1$, be a time change, i.e., a sequence of increasing stopping times, and define

$$X_t^k = X_{\tau_{k-1}+t} - X_{\tau_{k-1}}$$

with convention $\tau_0 = 0$ a.s. Subsequently, we denote by X^k the process (X_t^k) for $t \in [0, T]$ with $T > 0$ fixed, and regard (X^k) as a sequence of stochastic processes. Roughly, for each k, X^k is a stochastic process on $[0, T]$ defined from X in terms of increments relative to $X_{\tau_{k-1}}$. We assume

Assumption 5.1 Let (X^k) be strictly stationary and α-mixing.

Note that here we require the stationarity of (X^k) as a sequence in k, not that of X. We assume in general that X is nonstationary. The conditions in Assumption 5.1 are not very restrictive and, as we explain below, satisfied by a large class of stochastic processes used in the actual empirical researches and practical applications.

Many stochastic processes meet the conditions in Assumption 5.1 under the time change given by $\tau_k - \tau_{k-1} = \triangle\tau$ for some fixed $\triangle\tau > 0$, i.e., by some fixed increments in time for all k. For instance, they hold for all Lévy processes that have independent stationary increments. This simple time change makes the practical implementation of our methodologies prticularly easy and straightforward. If, in particular, we set $\triangle\tau = T$, then (X^k) become N-nonoverlapping subsets of X, all with the zero starting value. Strong markov processes like Brownian motion and Brownian motion with drift satisfy Assumption 5.1 with any time change τ_k if they have some nonzero minimal increment. This is so also for geometric Brownian motion, up to the logarithmic transformation. Moreover, Assumption 5.1 is met for all homogeneous α-mixing markov processes, including stationary diffusions such as OU process and Feller's SR process that are widely used in modelling interest rates. Indeed, all stationary homogeneous diffusions fulfill the assumption if we define τ_k sequentially to be the stopping time such that $\tau_k = \inf\{t \geq \tau_{k-1} | X_t = x\}$ for any $x \in \mathbb{R}$ with some nonzero minimal increment. It is well known that all stationary diffusions are α-mixing [see, e.g., Bosq (1998, p162)].

We now introduce the estimators for the spatial density and distribution function for the stochastic processes satisfying Assumption 5.1. Let $\hat{\ell}_k$ and $\hat{L}_k$, for $k = 1, \ldots, N$, be the estimators for the local time and integrated local time respectively that are introduced in (16) and (17), using discrete samples $(X_{i\Delta}^k)$ observed from X^k. Then we define

$$\hat{\lambda}_N(T, x) = \frac{1}{N} \sum_{k=1}^{N} \hat{\ell}_k(T, x) \tag{20}$$

$$\hat{\Lambda}_N(T, x) = \frac{1}{N} \sum_{k=1}^{N} \hat{L}_k(T, x). \tag{21}$$

The estimators here can be computed simply by averaging N-pieces of the local time and integrated local time estimators obtained for each of (X^k). Note that here we assume in particular that (X^k) have the same probability law for all k. For the time change (τ_k) given by $\tau_k - \tau_{k-1} = T$ for all k, we need observation on X over interval $[0, NT]$ to obtain $\hat{\lambda}_N(T,x)$ and $\hat{\Lambda}_N(T,x)$. For a general time change (τ_k), however, the required observation interval for X would be randomly given.

To develop the asymptotics for the spatial distributions of the processes with stationary increments, we need to introduce some additional technical conditions that are given in

Assumption 5.2 ℓ_k satisfies that $\sup_{t\leq T} |\ell_k(t,x) - \ell_k(t,y)| \leq C|x-y|^{1/2-\delta}$ for some $\delta > 0$ and random variable C, and that $\mathbb{E}|\ell_k(T,x)|^{2+\delta} < \infty$ for some $\delta > 0$.

The following theorems present the asymptotics for the estimators of the spatial density and distribution function that are introduced in (20) and (21) above for the processes with stationary increments. Here we denote by $\alpha(k)$ the mixing coefficient of (X^k), which is assumed to be α-mixing.

Theorem 5.1 *Suppose that* Assumptions 4.1, 5.1 *and* 5.2 *hold. We have*

(a) *Let h be chosen so that $\omega(\Delta)/h^{1+\delta} = o(1)$ with some $\delta > 0$ for all large N. Then, for all $x \in \mathbb{R}$, we have $\lambda_N(T,x) \to_{a.s.} \lambda(T,x)$ as $N \to \infty$. Furthermore, as $N \to \infty$,*

$$\int_{-\infty}^{\infty} \left| \hat{\lambda}_N(T,x) - \lambda(T,x) \right| dx \to_{a.s.} 0.$$

(b) *Let h be chosen so that $h^{1-\delta} = o(N^{-1})$ and $\omega(\Delta)/h^{1+\delta} = o(N^{-1/2})$ with some $\delta > 0$ for all large N, and let $\sum_{k=1}^{\infty} \alpha(k)^{\delta/(2+\delta)} < \infty$ for some $\delta > 0$. Then we have as $N \to \infty$*

$$\sqrt{N}\left(\hat{\lambda}_N(T,x) - \lambda(T,x)\right) \to_d \mathbf{N}(0, \sigma_S(T,x)),$$

where the asymptotic variance $\sigma_S(T,x)$ is given by

$$\begin{aligned} \sigma_S(T,x) &= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{k=1}^{N} [\ell_k(T,x) - \lambda(T,x)]\right)^2 \\ &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N}\sum_{j=1}^{N} \left(\lim_{\varepsilon\to 0} \frac{1}{4\varepsilon^2} \int_0^T\int_0^T \mathbb{E}\, I_t^\varepsilon(i,x) I_s^\varepsilon(j,x)\, dt\, ds\right) \end{aligned}$$

with $I_t^\varepsilon(k,x) = 1\{|X_t^k - x| < \varepsilon\} - \mathbb{P}\{|X_t^k - x| < \varepsilon\}$.

Theorem 5.2 *Suppose that* Assumption 5.1 *hold. We have*

(a) *Let $\omega(\Delta) = o(1)$ for all large N. Then we have as $N \to \infty$*

$$\sup_{x\in\mathbb{R}} \left| \hat{\Lambda}_N(T,x) - \Lambda(T,x) \right| \to_{a.s.} 0.$$

(b) *Let $\alpha(k) = O(k^{-9-\delta})$ for some $\delta > 0$, and let $\omega(\Delta) = o(N^{-1/2})$ for all large N. Then we have as $N \to \infty$*

$$\sqrt{N}\left(\hat{\Lambda}_N(T,\cdot) - \Lambda(T,\cdot)\right) \to_d U(T,\cdot),$$

where $U(T,\cdot)$ *is a mean zero Gaussian process with covariance kernel*

$$\begin{aligned}\mathbb{E}U(T,x)U(T,y) &= \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\left(\sum_{k=1}^{n}[L_k(T,x)-\Lambda(T,x)]\right)\left(\sum_{k=1}^{n}[L_k(T,y)-\Lambda(T,y)]\right)\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{N}\int_0^T\int_0^T \mathbb{E}J_t(i,x)J_s(j,y)\,dt\,ds\end{aligned}$$

with $J_t(k,w)=1\{X_t^k\le w\}-\mathbb{P}\{X_t^k\le w\}$.

Theorems 5.1 and 5.2 show that the spatial density estimator $\hat\lambda_N$ and the spatial distribution function estimator $\hat\Lambda_N$ in (20) and (21) are consistent. More precisely, it is established that $\hat\lambda_N(T,\cdot)$ is consistent a.s. and in L^1, and $\hat\Lambda_N(T,\cdot)$ is uniformly consistent a.s. The limit distributions for $\hat\lambda_N(T,x)$, $x\in\mathbb{R}$, as a sequence of random variables, and $\hat\Lambda_N(T,\cdot)$ as a sequence of random functions are also obtained.

It is clearly seen from Theorems 5.1 and 5.2 that $\hat\lambda_N(T,x)$, $x\in\mathbb{R}$, and $\hat\Lambda_N(T,\cdot)$ have the limit distributions that are dependent upon the probability law of the underlying stochastic process X in quite a complicated manner. There are several ways to obtain their limit distributions. If we know the complete law of X, then we may obviously use the standard simulation method. That is, we may generate from the law of X the random samples of size sufficiently large, and simulate the limit distributions of the their normalized statistics. As long as we know the spatial density $\lambda(T,x)$ at $x\in\mathbb{R}$, or the spatial distribution function $\Lambda(T,\cdot)$, we may use a bootstrap method such as the block bootstrap to find the limit distributions for $\hat\lambda_N(T,x)$, $x\in\mathbb{R}$, and $\hat\Lambda_N(T,\cdot)$, even if the complete law of X is unknown. For the bootstrap, we may simply regard as N serially correlated observations $\hat\ell_k(T,x)$, $x\in\mathbb{R}$, or $\hat L_k(T,\cdot)$ for $k=1,\ldots,N$ and draw samples from them to obtain the bootstrap samples of the statistics $\hat\lambda_N(T,x)$, $x\in\mathbb{R}$, and $\hat\Lambda_N(T,\cdot)$ introduced in (20) and (21).

In general, however, sub-sampling appears to be the method that is most readily available to obtain the limit distributions of $\hat\lambda_N(T,x)$, $x\in\mathbb{R}$, and $\hat\Lambda_N(T,\cdot)$, when the complete law of X is unknown. To compute their limit distributions by the sub-sampling method, we only need to observe that

$$\begin{aligned}\sqrt{N_s}\left(\hat\lambda_{N_s}(T,x)-\hat\lambda_N(T,x)\right) &\to_d \mathbb{N}(0,\sigma_S(T,x))\\ \sqrt{N_s}\left(\hat\Lambda_{N_s}(T,\cdot)-\hat\Lambda_N(T,\cdot)\right) &\to_d U(T,\cdot),\end{aligned}$$

where N_s is the size of sub-samples such that $N_s\to\infty$ and $N_s/N\to 0$, and $\sigma_S(T,x)$ and $U(T,\cdot)$ are introduced respectively in Theorems 5.1(b) and 5.2(b). This is quite obvious. The reader is referred to Politis, Romano and Wolf (1999) for the details and the general theory of the sub-sampling method. For the sub-sampling method here, we use $N-N_s+1$ number of sub-samples of size N_s to compute the asymptotic variance $\sigma_S(T,x)$ and the limit distribution given by $U(T,\cdot)$.

Now consider two stochastic processes X and Y. We let $Z=(X,Y)'$ be a vector process, for which we define $Z^k=(X^k,Y^k)'$ to be similarly as above. We assume

Assumption 5.3 Let (Z^k) be strictly stationary and α-mixing.

As before, we use $\alpha(k)$ to signify the mixing coefficient of (Z^k). The following corollary extends the distributional result in Theorem 5.2 to the multivariate case. Conformably as before, we define Λ^X and Λ^Y respectively to be the spatial distribution functions of X and Y, and $\hat{\Lambda}_N^X$ and $\hat{\Lambda}_N^Y$ to be their estimators defined as in (21) from the sample analogue estimators $\hat{L}_k^X$ and $\hat{L}_k^Y$ for the integrated local times of X^k and Y^k, $k = 1, \ldots, N$.

Corollary 5.3 *Suppose that* Assumption 5.3 *hold, and that* $\alpha(k) = O(k^{-9-\delta})$ *for some* $\delta > 0$ *and* $\omega(\Delta) = o(N^{-1/2})$ *for large* N. *Then we have as* $N \to \infty$

$$\sqrt{N} \begin{pmatrix} \hat{\Lambda}_N^X(T,\cdot) - \Lambda^X(T,\cdot) \\ \hat{\Lambda}_N^Y(T,\cdot) - \Lambda^Y(T,\cdot) \end{pmatrix} \to_d \begin{pmatrix} U^X(T,\cdot) \\ U^Y(T,\cdot) \end{pmatrix},$$

where $U^Z(T,\cdot) = (U^X(T,\cdot), U^Y(T,\cdot))'$ *is a mean zero vector Gaussian process with covariance kernel*

$$\mathbb{E}U^Z(T,x)U^Z(T,y)' = \lim_{N\to\infty} \frac{1}{N} \sum_{p=1}^{N} \sum_{q=1}^{N} \int_0^T \int_0^T \mathbb{E}J_t^Z(p,x)J_s^Z(q,y)'\,dt\,ds,$$

for which $J_t^Z(k,w) = (J_t^X(k,w), J_t^Y(k,w))'$ *with* $J_t^X(k,w)$ *and* $J_t^Y(k,w)$ *defined for the processes* X *and* Y *similarly as in* Theorem 5.2.

Obviously, the limit distribution $U^Z(T,\cdot)$ can be simulated if the probability law of Z is known. If $(\Lambda^X(T,\cdot), \Lambda^Y(T,\cdot))$ is known, the limit distribution can be computed using the bootstrap samples from the N observations $(\hat{L}_k^X(T,\cdot), \hat{L}_k^Y(T,\cdot))$, $k = 1, \ldots, N$, which are in general serially correlated. Otherwise, the limit distribution can be computed using the sub-sampling method precisely as in the univariate case.

Define the estimators for the d-spatial density and distribution function by

$$\begin{aligned} \hat{\lambda}_N^r(T,x) &= \frac{1}{N} \sum_{k=1}^{N} \hat{\ell}_k^r(T,x) \\ \hat{\Lambda}_N^r(T,x) &= \frac{1}{N} \sum_{k=1}^{N} \hat{L}_k^r(T,x) \end{aligned}$$

similarly as in (20) and (21), where $\hat{\ell}_k^r$ and $\hat{L}_k^r$ are the estimators of the d-local time and integrated d-local time in (18) and (19) obtained from X^k. We also denote by $\hat{\Lambda}_N^{r,X}$ and $\hat{\Lambda}_N^{r,Y}$ respectively the estimators for the d-spatial distribution functions of two stochastic processes X and Y.

Corollary 5.4 Theorems 5.1(a) *and* 5.2(a) *hold for* $(\hat{\lambda}_N^r, \lambda^r)$ *and* $(\hat{\Lambda}_N^r, \Lambda^r)$ *replacing respectively* $(\hat{\lambda}_N, \lambda)$ *and* $(\hat{\Lambda}_N, \Lambda)$. *Moreover,* Theorems 5.1(b) *and* 5.2(b) *hold for* $(\hat{\lambda}_N^r, \lambda^r)$ *and*

$(\hat{\Lambda}^r_N, \Lambda^r)$ in place of $(\hat{\lambda}_N, \lambda)$ and $(\hat{\Lambda}_N, \Lambda)$ with $\sigma_S(T,x)$ replaced by

$$\begin{aligned}\sigma^r_S(T,x) &= \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\left(\sum_{k=1}^n [\ell^r_k(T,x) - \lambda^r(T,x)]\right)^2 \\ &= \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \left(\lim_{\varepsilon\to 0}\frac{1}{4\varepsilon^2}\int_0^T\int_0^T \mathbb{E}\, I^{r,\varepsilon}_t(i,x) I^{r,\varepsilon}_s(j,x)\,dt\,ds\right),\end{aligned}$$

where $I^{r,\varepsilon}_t(k,x) = e^{-rt}\left(1\{|X^k_t - x| < \varepsilon\} - \mathbb{P}\{|X^k_t - x| < \varepsilon\}\right)$, and with $U(T,\cdot)$ replaced by $U^r(T,\cdot)$ which is a mean zero Gaussian process with covariance kernel

$$\begin{aligned}\mathbb{E}U^r(T,x)U^r(T,y) &= \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\left(\sum_{k=1}^n [L^r_k(T,x) - \Lambda^r(T,x)]\right)\left(\sum_{k=1}^n [L^r_k(T,y) - \Lambda^r(T,y)]\right) \\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^N\sum_{j=1}^N \int_0^T\int_0^T \mathbb{E}J^r_t(i,x)J^r_s(j,y)\,dt\,ds,\end{aligned}$$

where $J^r_t(k,w) = e^{-rt}\left(1\{X^k_t \le w\} - \mathbb{P}\{X^k_t \le w\}\right)$. Finally, Corollary 5.3 *holds for $(\hat{\Lambda}^{r,X}_N, \Lambda^{r,X})$ and $(\hat{\Lambda}^{r,Y}_N, \Lambda^{r,Y})$ instead of $(\hat{\Lambda}^X_N, \Lambda^X)$ and $(\hat{\Lambda}^Y_N, \Lambda^Y)$ with $U^Z(T,\cdot) = (U^X(T,\cdot), U^Y(T,\cdot))'$ substituted by $U^{r,Z}(T,\cdot) = (U^{r,X}(T,\cdot), U^{r,Y}(T,\cdot))'$, which is a mean zero vector Gaussian process with covariance kernel*

$$\mathbb{E}U^{r,Z}(T,x)U^{r,Z}(T,y)' = \lim_{N\to\infty}\frac{1}{N}\sum_{p=1}^N\sum_{q=1}^N\int_0^T\int_0^T \mathbb{E}J^{r,Z}_t(p,x)J^{r,Z}_s(q,y)'\,dt\,ds,$$

for which $J^{r,Z}_t(k,w) = (J^{r,X}_t(k,w), J^{r,Y}_t(k,w))'$ with $J^{r,X}_t(k,w)$ and $J^{r,Y}_t(k,w)$ defined for the processes X and Y similarly as above.

Therefore, the results in Theorems 5.1, 5.2 and Corollary 5.3 hold for the d-spatial density and distribution function and their corresponding estimators.

5.2 Markov Processes

Here we provide estimators of the spatial density and distribution function for markov processes. We assume that

Assumption 5.4 Let X be a homogeneous markov process, which has transition density $p(t,x,y)$ with respect to Lebesgue measure.

As is well known, the transition density completely specifies the probability law of a markov process.

Let $X_0 = x_0$ throughout this section. If X satisfies Assumption 5.4 and $p(t,x_0,x)$ is continuous in x for all $x \in \mathbb{R}$, then we have

$$\lambda(T,x) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T \mathbb{P}\{|X_t - x| < \varepsilon\}\,dt = \int_0^T p(t,x_0,x)\,dt. \tag{22}$$

Moreover, for any X satisfying Assumption 5.3, it follows that

$$\Lambda(T,x) = \int_0^T \mathbb{P}\{X_t \le x\}\,dt = \int_0^T \int_{-\infty}^x p(t,x_0,y)\,dy\,dt. \tag{23}$$

The spatial density and distribution function can therefore be estimated readily once the estimate for the transition density is obtained.

Now we explain how to estimate the transition density to facilitate our subsequent spatial analysis of time series. Let $M > 0$ be given. For each $x \in \mathbb{R}$, define κ_x to be such that

$$\ell(\kappa_x T, x) = M.$$

Therefore, $\kappa_x T$ is the observation interval making the value of local time uniform for all x, and be given by M. The transition density can be estimated by the usual kernel estimate, which is given by

$$\hat{p}_\Delta(\underline{\Delta}, x, y) = \frac{\displaystyle\sum_{i=1}^{n\kappa_x} K\left(\frac{X_{i\Delta}-x}{h}\right) K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)}{\displaystyle\underline{h}\sum_{i=1}^{n\kappa_x} K\left(\frac{X_{i\Delta}-x}{h}\right)},$$

where h and $\underline{h}$ are bandwidth parameters, and Δ and $\underline{\Delta}$ are intervals respectively at which the observations are made and the transition density is estimated. It turns out that it would yield better results if we estimate the transition density at intervals bigger than the length Δ of intervals on which the data are observed. We therefore assume that the transition density is estimated at the interval $\underline{\Delta}$ using the data observed at interval Δ, for which $\underline{\Delta} \ge \Delta$. On the other hand, it is preferred that the new bandwidth parameter $\underline{h}$ introduced to estimate the transition density is smaller than the original bandwidth parameter h, i.e., $\underline{h} \le h$.

Assumption 5.5 We assume (a) $p(t,\cdot,\cdot)$ are twice differentiable for all $t > 0$ with k-th order derivative $D^k p(t,\cdot,\cdot)$ satisfying $|D^k p(t,\cdot,\cdot)| \le c_k/t^{(1+k)/2}$ for some constants (c_k), $k = 0,1,2$, and (b) ℓ satisfies, for any stopping time τ, $\sup_{t\le\tau} |\ell(t,x) - \ell(t,y)| \le C_\tau |x-y|^{1/2-\delta}$ with some $\delta > 0$ and random variable C_τ.

The conditions in Assumption 5.5 are not very stringent. The conditions in (b) for the transition density hold for a wide variety of diffusion models including virtually all models used in practical applications. The interested reader is referred to, e.g., Friedman (1964, p251) and Florens-Zmirou (1993, p792) for the detailed discussions on these and other closely related conditions. Moreover, as shown in, e.g., Revuz and Yor (1994, pp227-228), the condition in (b) is met for a large class of continuous semimartingales whose bounded variation components do not explode at any finite time. The class, in particular, includes transient, as well as recurrent, processes such as Brownian motion with drift.

We now define X^Δ to be the markov process with transition density $\hat{p}_\Delta(t,x,y)$. Moreover, the underlying probability which renders the transition probability of X^Δ to be given by $\hat{p}_\Delta(t,x,y)$ is signified by $\mathbb{P}_\Delta$. The corresponding expectation will be denoted by $\mathbb{E}_\Delta$. Of course, the process X^Δ is defined only at discrete time. We would, however, regard it as a continuous process observed at discrete time intervals $\underline{\Delta}$.

We may now define $\hat\lambda_\Delta$ and $\hat\Lambda_\Delta$ to be the spatial density and distribution function of X^Δ using expectation $\mathbb{E}_\Delta$, i.e.,

$$\hat\lambda_\Delta(T,x) = \mathbb{E}_\Delta \ell_\Delta(T,x) \quad \text{and} \quad \hat\Lambda_\Delta(T,x) = \mathbb{E}_\Delta L_\Delta(T,x), \tag{24}$$

where ℓ_Δ and L_Δ are respectively the local time and integrated local time of X^Δ. Similarly as in (22) and (23), we have

$$\hat\lambda_\Delta(T,x) = \lim_{\varepsilon\to 0}\frac{1}{2\varepsilon}\int_0^T \mathbb{P}_\Delta\{|X_t^\Delta - x| < \varepsilon\}\,dt = \int_0^T \hat p_\Delta(t,x_0,x)\,dt \tag{25}$$

and

$$\hat\Lambda_\Delta(T,x) = \int_0^T \mathbb{P}_\Delta\{X_t^\Delta \le x\}\,dt = \int_0^T\int_{-\infty}^x \hat p_\Delta(t,x_0,y)\,dy\,dt. \tag{26}$$

The estimates for the spatial density $\hat\ell_\Delta$ and the spatial distribution function $\hat\Lambda_\Delta$ can be obtained from (24) through simulations or can be computed directly from (25) and (26) using the estimated transition density $\hat p_\Delta(t,x,y)$. In what follows, we let $\kappa_2 = \int_{-\infty}^{\infty} K^2(s)\,ds$.

Theorem 5.5 *Suppose that* Assumptions 4.1, 5.4 *and* 5.5 *hold. Let* $\omega(\Delta) = \Delta^{1/2-\delta}$ *for some* $\delta > 0$*, and suppose that we choose* $h = \Delta^{1/3}$, $\underline{\Delta} = \Delta^a$ *and* $\underline{h} = \Delta^b$ *for some constants* a *and* b *satisfying* $0 < a < 1/3$ *and* $a + 1/3 < b < -2a + 2/3$*. Then we have as* $\Delta \to 0$

$$\left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2}\left(\hat\lambda_\Delta(T,x) - \lambda(T,x)\right) \to_d \mathbb{N}(0,\sigma_K(T,x)),$$

where

$$\sigma_K(T,x) = \kappa_2^2\int_{-\infty}^{\infty} dw\left(\int_0^T dt\int_0^t ds\, p(s,x_0,w)p(t-s,w,x)\right)^2.$$

Moreover, we have as $\Delta \to 0$

$$\left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2}\left(\hat\Lambda_\Delta(T,\cdot) - \Lambda(T,\cdot)\right) \to_d V(T,\cdot),$$

where $V(T,\cdot)$ *is a mean zero Gaussian process with covariance kernel*

$$\begin{aligned}\mathbb{E}V(T,x)V(T,y) \;=\; &\kappa_2^2\int_{-\infty}^{\infty} dw\int_0^T dt\left(\int_0^t du\, p(u,x_0,w)\int_{-\infty}^x dz\, p(t-u,w,z)\right)\\ &\int_0^T ds\left(\int_0^s du\, p(u,x_0,w)\int_{-\infty}^y dz\, p(s-u,w,z)\right)\end{aligned}$$

for $x, y \in \mathbb{R}$.

The asymptotics in Theorem 5.5 for markov processes are obtained by setting various estimation parameters explicitly as functions of Δ and letting $\Delta \to 0$. This is in contrast with those in Theorem 5.1 for processes with stationary increments, where the asymptotics are developed by $N \to \infty$, requiring other estimation parameters to be given as functions of N. Our asymptotics here are developed in a way that we may best exploit the probability structure of each of these classes of models. Note that the former provides more information on its distribution as $N \to \infty$, i.e., as we observe more observations, while the probability law of the latter is determined by the transition density that we may estimate precisely only if $\Delta \to 0$, i.e., as we observe more frequently.[7]

All our conditions $h = \Delta^{1/3}$, $\underline{\Delta} = \Delta^a$ and $\underline{h} = \Delta^b$ in Theroem 5.5 may be defined up to constant multiples. In our asymptotics in Theorem 5.5, we may let M be either fixed or increasing as $\Delta \to 0$. Recall that M is the time span measured in the units of local time. Therefore, letting $M \to \infty$ along with $\Delta \to 0$ implies that we have observations over longer time span, as well as more frequently in any fixed time interval. Since the time span is usually limited by the availability of the data, we set $M = \Delta^{-\delta}$ for some small $\delta > 0$. If we set $a = \delta$ and $b = \delta + 1/3$ for $0 < \delta < 1/9$, then we have

$$\left(\frac{h\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} = \Delta^{-1/6+3\delta/2} \to \infty$$

as $\Delta \to 0$. Therefore, the estimators $\hat{\lambda}_\Delta$ and $\hat{\Lambda}_\Delta$ of the spatial density and spatial distribution function are consistent, with the rate of convergence given by $\Delta^{1/6-3\delta/2}$, if M is fixed. In general, it has the convergence rate $M^{-1/2}\Delta^{1/6-3\delta/2}$.

The limit distributions of $\hat{\lambda}_\Delta(T,x)$, $x \in \mathbb{R}$, and $\hat{\Lambda}_\Delta(T,\cdot)$ in Theorem 5.5 cannot be directly evaluated, unless we know the complete law of the underlying stochastic process X. We may, however, generally use a modified subsampling method to compute their limit distributions. For an observation interval Δ_s such that $\Delta_s \to 0$ and $\Delta_s/\Delta \to \infty$, we let $h_s = \Delta_s^{1/3}, \underline{\Delta}_s = \Delta_s^a$ and $\underline{h}_s = \Delta_s^b$ for the constants a and b introduced in Theorem 5.5. Also, we set $M_s = M$ fixed or $M_s \to \infty$ such that $M_s/M \to 0$. Then we have

$$\left(\frac{M_s h_s \underline{h}_s \underline{\Delta}_s^2}{\Delta_s}\right)^{1/2} \left(\hat{\lambda}_{\Delta_s}(T,x) - \hat{\lambda}_\Delta(T,x)\right) \to_d \mathbb{N}(0, \sigma_K(T,x))$$

$$\left(\frac{M_s h_s \underline{h}_s \underline{\Delta}_s^2}{\Delta_s}\right)^{1/2} \left(\hat{\Lambda}_{\Delta_s}(T,\cdot) - \hat{\Lambda}_\Delta(T,\cdot)\right) \to_d V(T,\cdot),$$

exactly as for the usual subsampling methods.

We now consider two processes X and Y, which are started at x_0 and y_0, respectively. As earlier, we let $Z = (X,Y)'$ and assume that

Assumption 5.6 Let Z be a homogeneous markov process, which has transition density with respect to Lebesgue measure.

[7]This is not true for stationary markov processes. We, however, consider primarily nonstationary markov processes, whose transition density is not consistently estimable unless $\Delta \to 0$.

Assumption 5.7 We assume (a) $p(t,\cdot,\cdot)$ are twice differentiable for all $t > 0$ with k-th order derivative $D^k p(t,\cdot,\cdot)$ satisfying $|D^k p(t,\cdot,\cdot)| \leq c_k/t^{1+k/2}$ for some constants (c_k), $k = 0, 1, 2$, and (b) ℓ^X and ℓ^Y satisfy, for any stopping time τ, $\sup_{t\leq\tau} |\ell^X(t,x) - \ell^X(t,y)| \leq C^X_\tau |x-y|^{1/2-\delta}$ and $\sup_{t\leq\tau} |\ell^Y(t,x) - \ell^Y(t,y)| \leq C^Y_\tau |x-y|^{1/2-\delta}$ with some $\delta > 0$ and random variables C^X_τ and C^Y_τ.

Corollary 5.6 *Suppose that* Assumptions 4.1, 5.6 *and* 5.7 *hold, and that the conditions in* Theorem 5.5 *are satisfied. Then we have as* $\Delta \to 0$

$$\left(\frac{Mh\underline{h}\Delta^2}{\Delta}\right)^{1/2} \begin{pmatrix} \hat{\Lambda}^X_\Delta(T,\cdot) - \Lambda^X(T,\cdot) \\ \hat{\Lambda}^Y_\Delta(T,\cdot) - \Lambda^Y(T,\cdot) \end{pmatrix} \to_d \begin{pmatrix} V^X(T,\cdot) \\ V^Y(T,\cdot) \end{pmatrix},$$

where $V^X(T,\cdot)$ *and* $V^Y(T,\cdot)$ *are independent Gaussian processes with covariance kernels given as in* Theorem 5.5 *for each of* X *and* Y*.*

The limiting distribution in Corollary 5.6 can be obtained in exactly the same manner as explained previously for the univariate case. Note that the simulation to compute the limiting distribution only requires the estimation of the univariate transition density. The estimation of the transition density for a vector process is unnecessary. As is well known, the transition density for a vector markov process is extremely difficult to precisely estimate and the estimation procedure is computationally quite burdensome.

The d-spatial density and distribution function can also be estimated using our method introduced above. Note that

$$\lambda^r(T,x) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_0^T e^{-rt}\mathbb{P}\{|X_t - x| < \varepsilon\}\, dt = \int_0^T e^{-rt} p(t,x_0,x)\, dt$$

and

$$\Lambda^r(T,x) = \int_0^T e^{-rt}\mathbb{P}\{X_t \leq x\}\, dt = \int_0^T e^{-rt} \int_{-\infty}^x p(t,x_0,y)\, dy\, dt,$$

which can be estimated respectively by

$$\hat{\lambda}^r_\Delta(T,x) = \lim_{\varepsilon\to 0} \frac{1}{2\varepsilon} \int_0^T e^{-rt}\mathbb{P}_\Delta\{|X^\Delta_t - x| < \varepsilon\}\, dt = \int_0^T e^{-rt} \hat{p}_\Delta(t,x_0,x)\, dt$$

and

$$\hat{\Lambda}^r_\Delta(T,x) = \int_0^T e^{-rt}\mathbb{P}_\Delta\{X^\Delta_t \leq x\}\, dt = \int_0^T e^{-rt} \int_{-\infty}^x \hat{p}_\Delta(t,x_0,y)\, dy\, dt,$$

or through the simulations based on samples generated by the estimated transition density $\hat{p}_\Delta(t,x,y)$. The asymptotic results for the d-spatial density and distribution function are given below in Corollary 5.7.

Corollary 5.7 Theorem 5.5 *holds for* $(\hat{\lambda}^r_\Delta, \lambda^r)$ *and* $(\hat{\Lambda}^r_\Delta, \Lambda^r)$ *with* $\sigma_K(T,x)$ *replaced by*

$$\sigma^r_K(T,x) = \kappa_2^2 \int_{-\infty}^{\infty} dw \left(\int_0^T dt\, e^{-rt} \int_0^t ds\, p(s,x_0,w) p(t-s,w,x) \right)^2,$$

and with $V(T,\cdot)$ *replaced by* $V^r(T,\cdot)$ *which is a mean zero Gaussian process with covariance kernel*

$$\begin{aligned}\mathbb{E}V^r(T,x)V^r(T,y) \;=\; \kappa_2^2 \int_{-\infty}^{\infty} dw \int_0^T dt\, e^{-rt} \left(\int_0^t du\, p(u,x_0,w) \int_{-\infty}^x dz\, p(t-u,w,z) \right) \\ \int_0^T ds\, e^{-rs} \left(\int_0^s du\, p(r,x_0,w) \int_{-\infty}^y dz\, p(s-u,w,z) \right).\end{aligned}$$

Moreover, Corollary 5.6 *holds for* $(\hat{\Lambda}^{r,X}_\Delta, \Lambda^{r,X})$ *and* $(\hat{\Lambda}^{r,Y}_\Delta, \Lambda^{r,Y})$ *instead of* $(\hat{\Lambda}^X_\Delta, \Lambda^X)$ *and* $(\hat{\Lambda}^Y_\Delta, \Lambda^Y)$ *with* $V^X(T,\cdot)$ *and* $V^Y(T,\cdot)$ *substituted respectively by* $V^{r,X}(T,\cdot)$ *and* $V^{r,Y}(T,\cdot)$.

The results in Corollary 5.7 for the d-spatial density and distribution function are comparable to those in Theorem 5.5 and Corollary 5.6 for the spatial density and distribution function.

6. Inferences in Spatial Distributions

In this section, we explore some immediate applications of the theories that are developed previously in the paper. Considered are four different applications that include forecast of spatial distribution, specification test in spatial domain, test of equality in spatial distribution and test of spatial dominance. They are presented below in sequel.

6.1 Forecast of Spatial Distribution

The spatial density of the underlying time series over a fixed time interval can be forecastable by the estimators we obtained in the previous section. Needless to say, they provide unbiased forecasts for the spatial density. Suppose that we are at time t and wish to obtain a forecast for the spatial density over the period $[t, t+T]$ given $X_t = x_t$. For the process with stationary increments, we may use

$$\frac{1}{N} \sum_{k=1}^N \hat{\ell}_k(T, x_t + x)$$

as an unbiased forecast, where $\hat{\lambda}_N$ is the estimator for the spatial density introduced in Section 5.1. If the underlying process also has the property of independent increments, this forecast has the minimum mean squared error and is therefore optimal. For the markov process,

$$\int_0^T \hat{p}_\Delta(t, x_t, x)\, dt$$

provides the optimal forecast in the sense of minimum mean squared error.

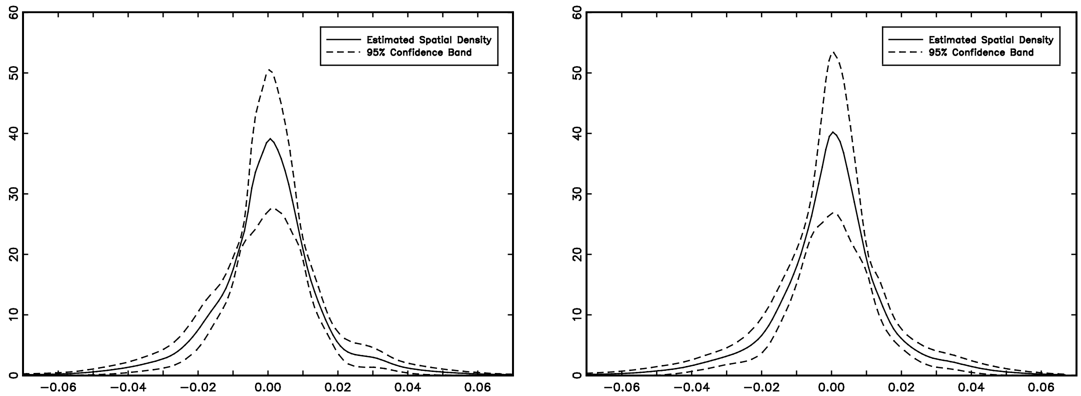

Figure 3: Predictive Spatial Densities: Logs of DJ and SP500

For given confidence level α, the confidence band for the expected local time can also be obtained using the theory developed in the previous section. That is, if we let $z_{\alpha/2}$ be the point that cuts off the $\alpha/2$-level tail probability from the standard normal distribution, the asymptotic α-level confidence interval for the true spatial density is given by

$$\left[\hat{\lambda}_N(T,x) - z_{\alpha/2}\sqrt{\frac{\sigma_S(T,x)}{N}},\ \hat{\lambda}_N(T,x) + z_{\alpha/2}\sqrt{\frac{\sigma_S(T,x)}{N}}\right]$$

for the process with stationary increments, and

$$\left[\hat{\lambda}_\Delta(T,x) - z_{\alpha/2}\sqrt{\frac{\sigma_K(T,x)\Delta}{Mh\underline{h}\underline{\Delta}^2}},\ \hat{\lambda}_\Delta(T,x) + z_{\alpha/2}\sqrt{\frac{\sigma_K(T,x)\Delta}{Mh\underline{h}\underline{\Delta}^2}}\right]$$

for the markov process, where $\sigma_S(T,x)$ and $\sigma_K(T,x)$ are defined respectively in Theorems 5.1 and 5.5.

In Figure 3, we present the predictive spatial densities over a week period for the log of DJ and SP500 stock indices. They are obtained as of 12/31/2004 to predict the spatial densities over the week of 01/03/2005 – 01/07/2005. The data collected at 30 minute intervals for the period of 01/07/2002 – 12/31/2004, comprising the total of 156 weeks, are used to estimate the spatial densities based on stationary increment models. The sampling time is simply set to have one week increment, so that the weekly observation units are non-overlapping. The point forecast is drawn in solid line, with 95% confidence bands given in dotted lines. As we explained in Section 3, the spatial densities provide, among other things, the aggregate VaR's. For instance, as of 12/31/2004, the aggregate VaR's over the week of 01/03/2005 – 01/07/2005 are given by 285.55 and 34.57 respectively for the (unlogged) DJ and SP500 stock indices at the 95% confidence level. On the day of 12/31/2004, we may therefore predict that the financial losses from the investments on the DJ and SP500 stock indices will not exceed 285.55 and 34.57, respectively, during the week of 01/03/2005 – 01/07/2005 with 95% aggregated weekly probability level. The values of the DJ and SP500 stock indices were 10785.22 and 1201.58, respectively, on the day of 12/31/2004.

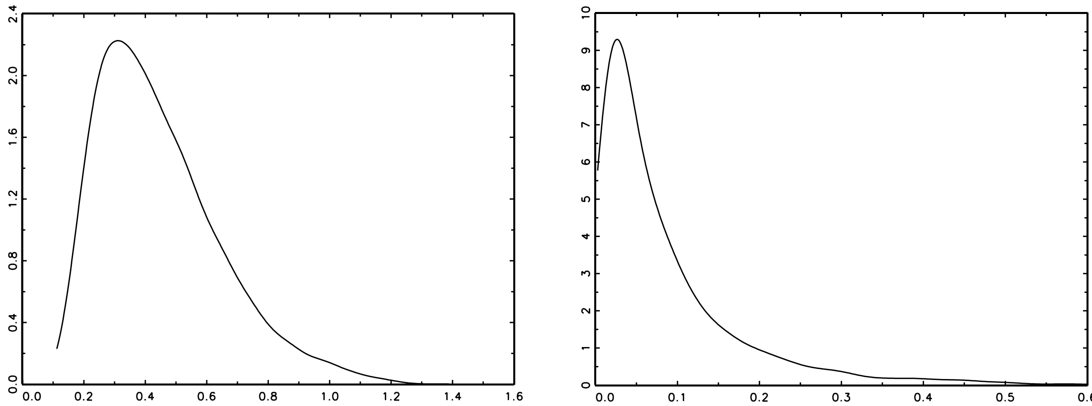

Figure 4: Limiting Distributions of $A_N^S(1)$ and $B_N^S(1)$ for Standard Brownian Motion

6.2 Specification Test in Spatial Domain

Using the theories developed in the previous section, we may test the hypothesis

$$\mathrm{H}_0 : \Lambda(T,\cdot) = \Lambda_0(T,\cdot),$$

where $\Lambda_0(T,\cdot)$ is a given spatial distribution function over time $[0,T]$. For instance, we may test whether the spatial distribution of a given stochastic process itself or any of its known transformation is the same as that of Brownian motion that is given in Section 2. The hypothesis can be tested using the Kolmogorov-Smirnov or the Cramér-von Mises type statistics that are commonly used to test the equality of the distributions of two random variables. We denote by $w(T,\cdot)$ the weight function used for the Cramér-von Mises type statistics. An obvious choice for the weight function in our context here would be $w(T,\cdot) = \lambda_0(T,\cdot)$, i.e., the spatial density under the null hypothesis.

In case of the process with stationary increment, we may use the test statistics

$$\begin{aligned} A_N^S(T) &= \sqrt{N} \sup_{x\in\mathbb{R}} \left| \hat{\Lambda}_N(T,x) - \Lambda_0(T,x) \right| \\ B_N^S(T) &= N \int_{-\infty}^{\infty} \left(\hat{\Lambda}_N(T,x) - \Lambda_0(T,x) \right)^2 w(T,x)\, dx, \end{aligned}$$

which have limiting null distributions given by

$$\begin{aligned} A_N^S(T) &\to_d \sup_{x\in\mathbb{R}} |U(T,x)| \\ B_N^S(T) &\to_d \int_{-\infty}^{\infty} U(T,x)^2 w(T,x)\, dx \end{aligned}$$

as one may easily deduce from Theorem 5.2 and the continuous mapping theorem.

For the markov process, we may similarly use the test statistics

$$A_\Delta^K(T) = \left(\frac{Mh\underline{h}\Delta^2}{\Delta} \right)^{1/2} \sup_{x\in\mathbb{R}} \left| \hat{\Lambda}_\Delta(T,x) - \Lambda_0(T,x) \right|$$

Table 1: Critical Values of $A_N^S(1)$ and $B_N^S(1)$

	Test Size		
Statistic	10%	5%	1%
$A_N^S(1)$	0.7237	0.8353	1.0411
$B_N^S(1)$	0.1990	0.2756	0.4532

$$B_\Delta^K(T) = \left(\frac{Mh\underline{h}\Delta^2}{\Delta}\right)\int_{-\infty}^{\infty}\left(\hat{\Lambda}_\Delta(T,x)-\Lambda_0(T,x)\right)^2 w(T,x)\,dx,$$

whose limiting null distributions are given by

$$\begin{aligned} A_\Delta^K(T) &\to_d \sup_{x\in\mathbb{R}}|V(T,x)| \\ B_\Delta^K(T) &\to_d \int_{-\infty}^{\infty} V(T,x)^2 w(T,x)\,dx, \end{aligned}$$

due to Theorem 5.5 and the continuous mapping theorem.

The limiting distributions of $A_N^S(T)$ and $B_N^S(T)$ are presented in Figure 4. Their asymptotic critical values are tabulated in Table 1. They were obtained by the simulation method based on 10,000 replications from observations made with $\Delta = 1/100$. We used the spatial density of the standard Brownian motion introduced in Section 2 as the weight function $w(T,\cdot)$ for $B_N^S(T)$). The time interval T is set to be unity. For the standard Brownian motion, we may easily deduce

$$\lambda(T,x) = \sqrt{T}\lambda\left(1,\frac{x}{\sqrt{T}}\right) \quad \text{and} \quad \Lambda(T,x) = T\Lambda\left(1,\frac{x}{\sqrt{T}}\right)$$

from the definition of the spatial density and distribution functions. Needless to say, the same relationships hold in the distributional sense for the estimated spatial density and distribution functions. Consequently, we have

$$A_N^S(T) =_d TA_N^S(1) \quad \text{and} \quad B_N^S(T) =_d T^3 B_N^S(1),$$

and therefore the limit distributions of $A_N^S(T)$ and $B_N^S(T)$ for general T may easily be obtained from those of $A_N^S(1)$ and $B_N^S(1)$. Of course, we may set $T = 1$ without loss of generality, unless we want to compare the spatial distributions over time intervals of different lengths.

The Brownian motion with non-unit variance can also be similarly dealt with, since any Brownian motion V, say, with variance σ^2, may be represented as $V = \sigma W$ using the standard Brownian motion W. Therefore, we have the distributional equivalence of

$$(V_t, 0 \le t \le T) \quad \text{and} \quad (W_t, 0 \le t \le \sigma^2 T)$$

in the spatial domain. The test statistics $A_N^S(T)$ and $B_N^S(T)$ constructed from V over time interval $[0,T]$ would thus have the limit distributions given by $\sigma^2 T$ and $\sigma^6 T^3$ multiples of

Table 2: Specification Test in Spatial Domain

Index	DJ		SP500	
Test Statistic	$A_N^s(1)$	$B_N^s(1)$	$A_N^s(1)$	$B_N^s(1)$
Test Value	1.1494	0.4396	1.0725	0.3928
P-Value	0.0030	0.0122	0.0080	0.0198

those provided here. For the nonstandard Brownian motion, we must therefore estimate the variance to implement our tests. The errors incurred from using the estimated variance would affect the null distributions of the tests up to a scalar factor. One way to make this error negligible is to estimate the variance using the samples that are substantially larger in magnitude than those used to compute the test statistics. Though we do not present the formal asymptotics for such procedures, it is obvious that the limit distributions would not change if we let the sample sizes for the variance estimators increase to infinity at a faster rate than those for the sample statistics.

As an application, we test whether the DJ and SP500 stock price indices are well specified in the spatial domain by the geometric Brownian motion. To implement the test, we log-transform the indices, remove the trend and adjust the mean and variance so that we may test whether the transformed series can be reasonably well fitted in the spatial domain by the standard Brownian motion. More precisely, the following steps are taken before we compare them with the standard Brownian motion in the spatial domain: First, we log-transform the stock indices and take the first differences. Second, estimate the mean and variance of the transformed first differences, and standardize them by subtracting the sample mean and dividing them by the sample standard deviation. The mean and standard deviation are estimated using considerably larger samples than the dataset we use to compute the statistics, so that the standardization does not affect the limiting distributions of the test statistics. Third, we integrate the standardized first differences to obtain the standardized stock indices in levels.

The results of our tests are summarized in Table 2. As before, we use the data collected at 30 minutes intervals for the period from 01/07/2002 to 12/31/2004 to obtain the total of 156 weekly sets of observations. The simple sampling time with one-week increment is used, so the obtained weekly units are non-overlapping. The mean and variance standardization are done using the same frequency data over the periods of 11/22/1996 – 12/31/2004 and 11/11/1997 – 12/31/2004 respectively for the DJ and SP500 stock indices. For both DJ and SP500 stock price indices, the specification of them as geometric Brownian motions in the spatial domain does not appear to be appropriate. Our tests $A_N^s(1)$ and $B_N^s(1)$ both reject the specification as Brownian motion rather strongly for the standardized log-transformations of the DJ and SP500 stock price indices. Recall that the spatial distributions of these indices determine the arbitrage-free prices of options written on them in a very definitive manner, as we have seen in Section 3. Of course, the rejection of the specification of these indices as geometric Brownian motions in the spatial domain implies that the celebrated Black-Scholes formula may not be very accurate in computing the arbitrage-free prices of options written on the DJ and SP500 stock price indices. Our test results here may thus explain the widely observed discrepancies between the Black-Scholes' and actual prices of the options on these indices.

6.3 Test of Equality in Spatial Distributions

Here we consider the test for the equality of spatial distributions. Let X and Y be two stochastic processes, and suppose we want to test the hypothesis

$$\mathrm{H}_0 : \Lambda^X(T,\cdot) = \Lambda^Y(T,\cdot),$$

where Λ^X and Λ^Y denote the spatial distribution functions of X and Y, respectively, as earlier. The processes X and Y need not be two distinct processes, i.e., they may be the single process observed in distinct time intervals. If we let X and Y be the process before and after a certain event, the test of the hypothesis may be used to analyze the effect of the event in the spatial domain. For instance, at the micro level, we may analyze the effect of stock splits by looking at the spatial distributions of the prices of a stock before and after it is splitted. We may also evaluate the effect of an economic policy or an intervention by comparing the spatial distributions before and after the introduction of the policy or intervention.

As for the specification test, we may use two types of statistics to test for the equality of spatial distributions: the Kolmogorov-Smirnov or the Cramér-von Mises type statistics. The former is based on the uniform distance, while the latter essentially looks at the L^2-distance, of the estimated expected integrated local times of X and Y. For the latter, we need to introduce a weight function and its estimate, which we denote respectively by $w(T,\cdot)$ and $\hat{w}(T,\cdot) = \hat{w}_N(T,\cdot)$ or $\hat{w}_\Delta(T,\cdot)$, respectively for processes with stationary increments or markov processes. We assume that $\int_{-\infty}^{\infty} |\hat{w}(T,x) - w(T,x)|\,dx \to_p 0$ as $N \to \infty$ or $\Delta \to 0$. Here we may use $\hat{w}(T,\cdot) = \hat{\lambda}(T,\cdot)$, where $\hat{\lambda} = \hat{\lambda}_N$ or $\hat{\lambda}_\Delta$ is the spatial density estimator using the samples from both X and Y. Note that $\lambda^X = \lambda^Y$ under the null hypothesis. Therefore, if we let λ be the common spatial density of X and Y, it follows from Theorem 5.1 or 5.5 that $\int_{-\infty}^{\infty} |\hat{\lambda}(T,x) - \lambda(T,x)|\,dx \to_p 0$. The required condition for the weight function would thus hold with $w(T,\cdot) = \lambda(T,\cdot)$ under the null hypothesis.

For the process with stationary increment, we may employ the statistic

$$\begin{aligned} P_N^S(T) &= \sqrt{N} \sup_{x\in\mathbb{R}} \left| \hat{\Lambda}_N^X(T,x) - \hat{\Lambda}_N^Y(T,x) \right| \\ Q_N^S(T) &= N \int_{-\infty}^{\infty} \left(\hat{\Lambda}_N^X(T,x) - \hat{\Lambda}_N^Y(T,x) \right)^2 \hat{w}_N(T,x)\,dx, \end{aligned}$$

which have limiting null distributions given by

$$\begin{aligned} P_N^S(T) &\to_d \sup_{x\in\mathbb{R}} |U^X(T,x) - U^Y(T,x)| \\ Q_N^S(T) &\to_d \int_{-\infty}^{\infty} (U^X(T,x) - U^Y(T,x))^2\, w(T,x)\,dx, \end{aligned}$$

Table 3: Test of Equality in Spatial Distributions

Test Statistic	$P_\Delta^K(T)$	$Q_\Delta^K(T)$
5% Critical Value	3.21×10^{-4}	3.72×10^{-8}
Test Value	9.03×10^{-5}	3.00×10^{-9}
P-Value	0.9367	0.9156

as follows readily from Corollary 5.3, together with the continuous mapping theorem.

For the markov process, we may similarly use the test statistics

$$P_\Delta^K(T) = \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \sup_{x\in\mathbb{R}} \left| \hat{\Lambda}_\Delta^X(T,x) - \hat{\Lambda}_\Delta^Y(T,x)\right|$$
$$Q_\Delta^K(T) = \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right) \int_{-\infty}^{\infty} \left(\hat{\Lambda}_\Delta^X(T,x) - \hat{\Lambda}_\Delta^Y(T,x)\right)^2 \hat{w}_\Delta(T,x)\,dx,$$

whose limiting null distributions are given by

$$P_\Delta^K(T) \to_d \sup_{x\in\mathbb{R}} |V^X(T,x) - V^Y(T,x)|$$
$$Q_\Delta^K(T) \to_d \int_{-\infty}^{\infty} (V^X(T,x) - V^Y(T,x))^2 w(T,x)\,dx,$$

due to Corollary 5.6 and the continuous mapping theorem.

As an illustrative example, we test for the equality of the spatial distributions of the returns from the DJ and SP500 stock indices. The returns were obtained by taking log differences of their levels. The tests are based on the statistics $P_\Delta^K(T)$ and $Q_\Delta^K(T)$, and their results are presented in Table 3. To implement the tests, we use $h = 0.004\Delta^{1/3}$, $\underline{\Delta} = 0.04\Delta^{1/18}$ and $\underline{h} = 0.0015\Delta^{7/18}$. These choices of h, $\underline{\Delta}$ and $\underline{h}$ satisfy the conditions in Theorem 5.5, and yield the convergence rate $M^{-1/2}\Delta^{1/12}$ for the estimators of the spatial density and distribution functions. The constant values are obtained through cross-validations.[8] The test values are computed using the 30-minutes data, as in our earlier examples, and for the level $M = 2,000$ of the local time, as of the closing time on 12/31/2004. The time span of the data used in the actual computation differs for each spatial point. Over the interval $[-0.001545, 0.001545]$, the length of sample is chosen differently for each of the spatial points so that the value of the local time estimate is uniformly given by 2,000. This requires data with time spans varying from 01/05/2004 – 12/31/2004 to 12/16/2004 – 12/31/2004. We simply use all the available data for the spatial point outside this interval.

The critical values are computed using a subsampling method, as explained before in Section 5. To obtain the test values based on the entire sample, we set $M = 4,000$ and $\Delta = 5$ minutes. Over the range $[-0.001727, 0.001667]$ in the spatial domain, the level $M = 4,000$ of local time is attained if we use the data spanning about a year, i.e., 01/06/2003 – 12/31/2004. Exactly as we did to compute the test values above, we go back from

[8]The values used here were obtained through a preliminary run. They will be replaced later by more accurate values obtained from more extensive and larger scale simulations.

12/31/2004 as much as needed into the past for each spatial point so that we reach this preset level of local time. Outside the interval $[-0.001727, 0.001667]$, we simply use all the available samples. The subsamples are obtained with the specification $M_s = 2,000$ and $\Delta_s = 30$ minutes through similar procedures. The subsamples are generated daily starting from 12/31/2004 and moving backward. The total number of 237 subsamples are generated to calculate the critical values. For each of the subsamples, we make the estimated value of the local time uniform at the level of 2,000 over $[-0.001545, 0.001545]$, the same interval that we used to compute the test values. Outside the interval, we use all the available samples as before for all the test values obtained from subsamples.

Our test results do not reject the equality of the spatial distributions of the DJ and SP500 stock returns. Though they often show somewhat distinct patterns of behaviors, our tests show that their spatial distributions are not significantly different from each other. The reported results appear to be pretty robust. Though we do not report the details here, we have the same results under a wide variety of possible specifications of the parameters of our nonparametric methods, and also across different sample periods and sampling frequencies.

6.4 Test of Spatial Dominance

As above, we consider two processes X and Y, and denote their spatial distribution functions by Λ^X and Λ^Y. Define

$$\begin{aligned} \delta(T) &= \sup_{x\in\mathbb{R}} \left(\Lambda^X(T,x) - \Lambda^Y(T,x)\right) \\ \delta^r(T) &= \sup_{x\in\mathbb{R}} \left(\Lambda^{r,X}(T,x) - \Lambda^{r,Y}(T,x)\right). \end{aligned}$$

The hypothesis of interest can now be stated as

$$\mathrm{H}_0 : \delta(T) \leq 0 \quad \text{or} \quad \mathrm{H}_0^r : \delta^r(T) \leq 0,$$

which is tested against the alternative hypothesis $\mathrm{H}_1 : \delta(T) > 0$ or $\mathrm{H}_1^r : \delta^r(T) > 0$. The test of H_0 against H_1 or H_0^r against H_1^r will be referred to as the test of spatial dominance in the paper.

Our notion of spatial dominance generalizes that of stochastic dominance. The latter is valid only for stationary processes, while the former is applicable for nonstationary processes as well. There are several different concepts of stochastic dominance, among which we only consider in the paper what is known as the first order stochastic dominance. See, e.g., Linton, Maasoumi and Whang (2003) for more details.

To test for the spatial dominance of X over Y without time discount, we may use the statistic defined by

$$D_N^S(T) = \sqrt{N} \sup_{x\in\mathbb{R}} \left(\hat{\Lambda}_N^X(T,x) - \hat{\Lambda}_N^Y(T,x)\right)$$

in the case of the process with stationary increments, and

$$D_\Delta^K(T) = \left(\frac{Nh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \sup_{x\in\mathbb{R}} \left(\hat{\Lambda}_\Delta^X(T,x) - \hat{\Lambda}_\Delta^Y(T,x)\right)$$

Table 4: Test of Spatial Dominance

Test Statistic	$D^K_\Delta(1)$	
Hypothesis	$\Lambda^{\text{DJ}} \leq \Lambda^{\text{SP500}}$	$\Lambda^{\text{SP500}} \leq \Lambda^{\text{DJ}}$
5% Critical Value	2.73×10^{-4}	3.06×10^{-4}
Test Value	9.02×10^{-5}	2.63×10^{-5}
P-Value	0.60759	0.89029

for the markov process. Their limiting null distributions are given by

$$D^S_N(T) \to_d \sup_{x\in\mathbb{R}} \left(U^X(T,x) - U^Y(T,x)\right)$$
$$D^K_\Delta(T) \to_d \sup_{x\in\mathbb{R}} \left(V^X(T,x) - V^Y(T,x)\right)$$

if $\delta(T) = 0$, and $D^S_N(T), D^K_\Delta(T) \to_p -\infty$ if $\delta(T) < 0$.

For the test of the spatial dominance of X over Y with time discount, we may similarly use the statistics given by

$$D^{r,S}_N(T) = \sqrt{N} \sup_{x\in\mathbb{R}} \left(\hat{\Lambda}^{r,X}_N(T,x) - \hat{\Lambda}^{r,Y}_N(T,x)\right)$$
$$D^{r,K}_\Delta(T) = \left(\frac{Mh\underline{h}\Delta^2}{\Delta}\right)^{1/2} \sup_{x\in\mathbb{R}} \left(\hat{\Lambda}^{r,X}_\Delta(T,x) - \hat{\Lambda}^{r,Y}_\Delta(T,x)\right)$$

respectively for the processes with stationary increments and the markov processes. They have the limiting distributions

$$D^{r,S}_N(T) \to_d \sup_{x\in\mathbb{R}} \left(U^{r,X}(T,x) - U^{r,Y}(T,x)\right)$$
$$D^{r,K}_\Delta(T) \to_d \sup_{x\in\mathbb{R}} \left(V^{r,X}(T,x) - V^{r,Y}(T,x)\right)$$

if $\delta^r(T) = 0$, and $D^{r,S}_N(T), D^{r,K}_\Delta(T) \to_p -\infty$ if $\delta^r(T) < 0$.

For an empirical application, we consider the returns from the DJ and SP500 stock price indices using the test based on $D^K_\Delta(1)$. The test results are presented in Table 4. The results are based on the same dataset used in our earlier illustration on the test of equality in spatial distributions. The choices of $h, \underline{\Delta}$ and $\underline{h}$ are also made exactly as we specify there. Moreover, the same subsampling method is employed to generate the critical values of the test. The test is done on 12/31/2004 for the week of 01/03/2005 – 01/07/2005. On the closing of the day 12/31/2004, the values of the returns are −0.00002689 and 0.00022281 respectively for the DJ and SP500 indices. The return from the SP500 index is substantially higher than that from the DJ index. The test results are clear and unambiguous: None of the returns from the DJ and SP500 indices spatially dominates the other. This implies that, as of 12/31/2004, none of the two indices is predicted to yield a higher level of the expected utility over the week of 01/03/2005 – 01/07/2005 for anybody with a nondecreasing utility function. This is so even though there are nonnegligible differences in their returns at the time when we evaluate their perspective weekly spatial profiles.

7. Concluding Remarks

In this paper, we develop a new methodology which is called the spatial analysis of time series. The spatial analysis allows us to investigate a time series along the spatial axis, i.e., the axis for its realized values, instead of the usual time axis. Our methodology exploits the fact that the distribution of a time series along the spatial axis is given by the local time of the stochastic process that generates the given time series. In particular, we define the spatial distribution of a time series to be the distribution given by the expected local time of the underlying stochastic process. Subsequently, we show that it is the spatial distribution, not the underlying stochastic process itself, which determines the expected future utilities that an economic time series generates. It can therefore be easily seen that the spatial analysis is the most essential part of the empirical analysis in any problem involving dynamic decision making based on the expected utility. We provide some of such examples for the purpose of illustrations.

A variety of statistical methods are introduced in the paper to facilitate the analysis of the spatial distribution. We provide all the essential procedures and relevant asymptotic theories that are required for the implementation of our methodology in practical applications. However, we still have many open questions, especially on the optimal choices of various parameters in our nonparametric procedures. The spatial analysis is primarily for the time series that do not have time invariant stationary distributions. Moreover, our framework is very general and imposes only minimal assumptions on the structure of the underlying stochastic process. This is why we cannot rely much on the existing literature, where the time invariant stationarity is routinely assumed and more structural assumptions are imposed. In particular, not much statistical theory is available in the literature to effectively deal with nonstationarity at the level of generality required in the paper. A further new development of the methodology that is valid for the general nonstationary stochastic process is therefore necessary to make the spatial analysis more reliable and more easily implemented in practice.

Appendix: Mathematical Proofs

Appendix A: Useful Lemmas and Their Proofs

Lemma A1 *We have*

$$\int_0^T e^{-rt}u(X_t)\,dt = \int_{-\infty}^{\infty} u(x)\ell^r(T,x)\,dx$$

for any nonnegative Borel-measureable function u on $\mathbb{R}$.

Proof of Lemma A1 It follows from the extended version of the occupation times formula in, e.g., Revuz and Yor (1994, Exercise 1.15, p.222) that

$$\int_0^T e^{-rt}u(X_t)\,dt \;=\; \int_{-\infty}^{\infty} dx \int_0^T e^{-rt}u(x)\,\ell(dt,x)$$

$$= \int_{-\infty}^{\infty} u(x)\ell^r(T,x)\,dx,$$

as was to be shown. □

Lemma A2 *If $\ell(T,\cdot)$ is continuous a.s., then so is $\ell^r(T,\cdot)$.*

Proof of Lemma A2 Consider a realization of X, and let $\mathcal{X}$ be the support of $\ell(T,\cdot)$. Note that $\mathcal{X}$ is compact a.s., since X has continuous sample path a.s. Moreover, $\ell(t,x)$ is continuous in both $t \in [0,T]$ and $x \in \mathcal{X}$, and being so on a compact domain, ℓ is uniformly continuous on $[0,T]\times\mathcal{X}$. Therefore, we have

$$\sup_{0\le t\le T} |\ell(t,x) - \ell(t,y)| \to 0$$

as $|x-y| \to 0$. We now note that

$$\begin{aligned} |\ell^r(T,x) - \ell^r(T,y)| &\le \left| \int_0^T e^{-rt}\ell(dt,x) - \int_0^T e^{-rt}\ell(dt,y) \right| \\ &\le \int_0^T |\ell(dt,x) - \ell(dt,y)| \\ &\le 2 \sup_{0\le t\le T} |\ell(t,x) - \ell(t,y)| \qquad (27) \end{aligned}$$

which, as shown above, goes to zero as $|x-y| \to 0$. This completes the proof. □

Lemma A3 If $\ell(t,\cdot)$ is Hölder continuous of order p uniformly in t on $[0,T]$, then $\ell^r(T,\cdot)$ is Hölder continuous of order p.

Proof of Lemma A3 The stated result follows immediately from (27). □

Appendix B: Proofs of the Main Results

Proof of Lemma 2.1 The stated result can be easily derived by taking expectation to both sides of the occupation time formula in (2) and applying Fubini's theorem. □

Proof of Corollary 2.2 The stated result follows immediately from Lemma A1, precisely as in the proof of Lemma 2.1. □

Proof of Theorem 4.1 We let

$$\tilde{\ell}(T,x) = \frac{1}{h}\int_0^T K\left(\frac{X_t - x}{h}\right) dt,$$

and write

$$\hat{\ell}(T,x) - \ell(T,x) = \left[\hat{\ell}(T,x) - \tilde{\ell}(T,x)\right] + \left[\tilde{\ell}(T,x) - \ell(T,x)\right]. \qquad (28)$$

To prove the first part, we will show that both terms in (28) become negligible as $\Delta \to 0$, under the given conditions.

To do so, we first establish that

$$\sup_{x\in\mathbb{R}} \left|\hat{\ell}(T,x) - \tilde{\ell}(T,x)\right| \leq \frac{\omega(\Delta)}{h^{1+\delta}} \left[c_1 + c_2 \left(\sup_{x\in\mathbb{R}} \ell(T,x)\right)\right] \tag{29}$$

for all $\omega(\Delta)/h^{1+\delta}$ sufficiently small, where c_1 and c_2 are constants depending only upon δ, T and K. We write

$$\hat{\ell}(T,x) = \tilde{\ell}(T,x) + R(T,x),$$

where

$$R(T,x) = \frac{1}{h}\sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left[K\left(\frac{X_{i\Delta}-x}{h}\right) - K\left(\frac{X_t - x}{h}\right)\right] dt,$$

and consider the a-th order Taylor expansion given by

$$\begin{aligned} K\left(\frac{X_{i\Delta}-x}{h}\right) - K\left(\frac{X_t-x}{h}\right) &= \left(\frac{X_{i\Delta}-X_t}{h}\right) K_1\left(\frac{X_t-x}{h}\right) \\ &+ \frac{1}{2}\left(\frac{X_{i\Delta}-X_t}{h}\right) K_2\left(\frac{X_t-x}{h}\right) + \cdots + \frac{1}{a!}\left(\frac{X_{i\Delta}-X_t}{h}\right)^a K_a\left(\frac{X_{s_i}-x}{h}\right) \end{aligned}$$

with some s_i's such that $t \leq s_i \leq i\Delta$. It follows directly from the definition of $R(T,x)$ and the Taylor expansion that

$$\begin{aligned} |R(T,x)| \leq\; & \frac{\omega(\Delta)}{h}\frac{1}{h}\int_0^T \left|K_1\left(\frac{X_t-x}{h}\right)\right| dt + \left(\frac{\omega(\Delta)}{h}\right)^2 \frac{1}{h}\int_0^T \left|K_2\left(\frac{X_t-x}{h}\right)\right| dt \\ & + \cdots + \frac{T}{a!}\frac{1}{h}\left(\frac{\omega(\Delta)}{h}\right)^a \left(\sup_{s\in\mathbb{R}} |K_a(s)|\right), \end{aligned} \tag{30}$$

since in particular

$$\left|\frac{X_{i\Delta}-X_t}{h}\right| \leq \frac{\omega(\Delta)}{h}$$

for all $(i-1)\Delta \leq t \leq i\Delta$.

Note that we have for all $b = 1, \ldots, a-1$

$$\begin{aligned} \frac{1}{h}\int_0^T \left|K_b\left(\frac{X_t-x}{h}\right)\right| dt &= \frac{1}{h}\int_{-\infty}^{\infty} \left|K_b\left(\frac{s-x}{h}\right)\right| \ell(T,s)\, dt \\ &= \int_{-\infty}^{\infty} |K_b(s)| \ell(T, x+hs)\, dt \\ &\leq \left(\sup_{x\in\mathbb{R}} \ell(T,x)\right) \int_{-\infty}^{\infty} |K_b(s)| dt \end{aligned}$$

for all $x \in \mathbb{R}$, and that

$$\left(\frac{\omega(\Delta)}{h}\right)^b < \left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)^b \leq \frac{\omega(\Delta)}{h^{1+\delta}}$$

for all $b = 1, \ldots, a-1$. Moreover, if we set a sufficiently large so that $1/\delta < a$, then we have

$$\frac{\omega(\Delta)^a}{h^{a+1}} = \left(\frac{\omega(\Delta)}{h^{1+1/a}}\right)^a < \left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)^a < \frac{\omega(\Delta)}{h^{1+\delta}}.$$

Consequently, if we let

$$c_1 = \frac{T}{a!} \sup_{s \in \mathbb{R}} |K_a(s)| \quad \text{and} \quad c_2 = \max_{1 \le b \le a-1} \int_{-\infty}^{\infty} |K_b(s)| \, ds,$$

the result in (29) follows immediately from (30).

Now, by the successive applications of the occupation times formula, change-of-variables and dominated convergence, we may easily deduce that

$$\begin{aligned} \tilde{\ell}(T,x) &= \frac{1}{h} \int_0^T K\left(\frac{X_t - x}{h}\right) dt \\ &= \frac{1}{h} \int_{-\infty}^{\infty} K\left(\frac{s-x}{h}\right) \ell(T,s) \, ds \\ &= \int_{-\infty}^{\infty} K(s) \ell(T, x+hs) \, ds \to_{a.s.} \ell(T,x), \end{aligned} \tag{31}$$

as $h \to 0$. This, together with (28) and (29), would imply

$$\hat{\ell}(T,x) \to_{a.s.} \ell(T,x) \tag{32}$$

for all $x \in \mathbb{R}$, under the given condition.

The stated result in the second part follows readily from (32), as in the proof of the so-called Scheffé's theorem [see, e.g., Serfling (1980, Theorem C, p17)], since in particular we have

$$\int_{-\infty}^{\infty} \hat{\ell}(T,x) \, dx = \int_{-\infty}^{\infty} \ell(T,x) \, dx = T.$$

To prove the second part, we let

$$\delta(T,x) = \left[\ell(T,x) - \hat{\ell}(T,x)\right] 1 \left\{\ell(T,x) \ge \hat{\ell}(T,x)\right\}.$$

and note that $\delta(T,x) \le \ell(T,x)$ for all $x \in \mathbb{R}$, and $\delta(T,x) \to_{a.s.} 0$ for each $x \in \mathbb{R}$ as shown in (32). As a result, it follows that

$$\int_{-\infty}^{\infty} \left|\hat{\ell}(T,x) - \ell(T,x)\right| dx = 2 \int_{-\infty}^{\infty} \delta(T,x) \, dx \to_{a.s.} 0,$$

due to dominated convergence. The proof is therefore complete. □

Proof of Theorem 4.2 Let

$$R(T,x) = \hat{L}(T,x) - L(T,x)$$
$$= \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \Big(1\{X_{i\Delta} \le x\} - 1\{X_t \le x\} \Big)\, dt \tag{33}$$

for $x \in \mathbb{R}$. Then we have

$$\int_{(i-1)\Delta}^{i\Delta} \Big| 1\{X_{i\Delta} \le x\} - 1\{X_t \le x\} \Big|\, dt \le \int_{(i-1)\Delta}^{i\Delta} 1\{|X_t - x| \le \omega(\Delta)\}\, dt, \tag{34}$$

and therefore,

$$|R(T,x)| \le \int_0^T 1\{|X_t - x| \le \omega(\Delta)\}\, dt \tag{35}$$

for all $x \in \mathbb{R}$. Moreover, it follows from the successive applications of the occupation times formula and change-of-variables that

$$\frac{1}{\omega(\Delta)} \int_0^T 1\{|X_t - x| \le \omega(\Delta)\}\, dt = \frac{1}{\omega(\Delta)} \int_{-\infty}^{\infty} 1\{|s - x| \le \omega(\Delta)\}\, \ell(T,s)\, ds$$
$$= \int_{-\infty}^{\infty} 1\{|s| \le 1\} \ell(T, x + s\omega(\Delta))\, ds$$
$$\le 2 \left(\sup_{x\in\mathbb{R}} \ell(T,x) \right) \tag{36}$$

for all $x \in \mathbb{R}$. Now we may easily deduce from (33), (35) and (36) that

$$\sup_{x\in\mathbb{R}} \Big| \hat{L}(T,x) - L(T,x) \Big| \le 2\omega(\Delta) \left(\sup_{x\in\mathbb{R}} \ell(T,x) \right), \tag{37}$$

from which the stated result follows immediately. □

Proof of Corollary 4.3 As in the proof of Theorem 4.1, we define

$$\tilde{\ell}^r(T,x) = \frac{1}{h} \int_0^T e^{-rt} K\left(\frac{X_t - x}{h} \right) dt,$$

and write

$$\hat{\ell}^r(T,x) = \tilde{\ell}^r(T,x) + R(T,x), \tag{38}$$

where

$$R(T,x) = \frac{1}{h} \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} \left[e^{-ri\Delta} K\left(\frac{X_{i\Delta} - x}{h} \right) - e^{-rt} K\left(\frac{X_t - x}{h} \right) \right] dt.$$

Subsequently, we let

$$R(T,x) = R_1(T,x) + R_2(T,x), \tag{39}$$

where

$$R_1(T,x) \;=\; \frac{1}{h}\sum_{i=1}^{n} K\!\left(\frac{X_{i\Delta}-x}{h}\right)\int_{(i-1)\Delta}^{i\Delta}\left(e^{-ri\Delta}-e^{-rt}\right)dt$$

$$R_2(T,x) \;=\; \frac{1}{h}\sum_{i=1}^{n}\int_{(i-1)\Delta}^{i\Delta} e^{-rt}\left[K\!\left(\frac{X_{i\Delta}-x}{h}\right)-K\!\left(\frac{X_t-x}{h}\right)\right]dt.$$

Below we will show that both $R_1(T,x)$ and $R_2(T,x)$ become negligible uniformly in $x \in \mathbb{R}$.

For $R_1(T,x)$, we note that

$$\max_{1\le i\le n}\left|\int_{(i-1)\Delta}^{i\Delta}\left(e^{-ri\Delta}-e^{-rt}\right)dt\right| \;\le\; \Delta \max_{1\le i\le n}\left(e^{-r(i-1)\Delta}-e^{-ri\Delta}\right)$$
$$= \Delta \max_{1\le i\le n}\int_{(i-1)\Delta}^{i\Delta} re^{-rt}dt \le r\Delta^2, \tag{40}$$

and therefore,

$$|R_1(T,x)| \le r\Delta\,\hat{\ell}(T,x) \le r\Delta\left(\tilde{\ell}(T,x)+\sup_{x\in\mathbb{R}}\left|\hat{\ell}(T,x)-\tilde{\ell}(T,x)\right|\right)$$

uniformly in $x \in \mathbb{R}$. However, we have as follows from (31)

$$\tilde{\ell}(T,x) \le \sup_{x\in\mathbb{R}} \ell(T,x)$$

uniformly in $x \in \mathbb{R}$, which together with (29) yields

$$|R_1(T,x)| \le c_1 r\Delta\frac{\omega(\Delta)}{h^{1+\delta}} + \left(r\Delta + c_2 r\Delta\frac{\omega(\Delta)}{h^{1+\delta}}\right)\left(\sup_{x\in\mathbb{R}}\ell(T,x)\right) \tag{41}$$

uniformly in $x \in \mathbb{R}$.

For $R_2(T,x)$, we let K_q be the q-th order derivative of K, and note that

$$\frac{1}{h}\int_0^T e^{-rt}\left|K_q\!\left(\frac{X_t-x}{h}\right)\right|dt \;=\; \frac{1}{h}\int_{-\infty}^{\infty}\left|K_q\!\left(\frac{s-x}{h}\right)\right|\ell^r(T,s)\,dt$$
$$= \int_{-\infty}^{\infty}|K_q(s)|\ell^r(T,x+hs)\,dt$$
$$\le \left(\sup_{x\in\mathbb{R}}\ell^r(T,x)\right)\int_{-\infty}^{\infty}|K_q(s)|dt$$

for all $x \in \mathbb{R}$. Therefore, we may deduce

$$|R_2(T,x)| \le \frac{\omega(\Delta)}{h^{1+\delta}}\left[c_1 + c_2\left(\sup_{x\in\mathbb{R}}\ell(T,x)\right)\right] \tag{42}$$

uniformly in $x \in \mathbb{R}$, as (29) in the proof of Theorem 4.1. Consequently, we have

$$\sup_{x\in\mathbb{R}}\left|\hat{\ell}^r(T,x)-\tilde{\ell}^r(T,x)\right| \le c_1(1+r)\Delta\frac{\omega(\Delta)}{h^{1+\delta}} + \left(r\Delta + c_2(1+r)\Delta\frac{\omega(\Delta)}{h^{1+\delta}}\right)\left(\sup_{x\in\mathbb{R}}\ell(T,x)\right) \tag{43}$$

from (38), (39), (41) and (42).

We now show that

$$\tilde{\ell}^r(T,x) \to_{a.s.} \ell^r(T,x) \tag{44}$$

as $h \to 0$, which together with (43) would complete our proof here. To deduce (44), we simply apply the extended occupation times formula, change-of-variables and the continuity of the d-local time in Lemma A2, and obtain

$$\begin{aligned} \frac{1}{h}\int_0^T e^{-rt} K\left(\frac{X_t - x}{h}\right) dt &= \frac{1}{h}\int_{-\infty}^{\infty} K\left(\frac{s-x}{h}\right) \ell^r(T,s)\, ds \\ &= \int_{-\infty}^{\infty} K(s)\ell^r(T, x+hs)\, ds \\ &\to_{a.s.} \ell^r(T,x) \end{aligned}$$

as $h \to 0$.

The proof for the L^1-convergence of the d-local time is essentially identical to that of the local time given in the proof of Theorem 4.1. We first note that

$$\int_{-\infty}^{\infty} \hat{\ell}^r(T,x)\, dx = \Delta \sum_{i=1}^n e^{-ri\Delta} = \int_0^T e^{-rt}\, dt + O(\Delta),$$

and that

$$\int_{-\infty}^{\infty} \ell^r(T,x)\, dx = \int_{-\infty}^{\infty}\int_0^T e^{-rt}\ell(dt,x)\, dx = \int_0^T e^{-rt} dt.$$

Therefore, if we let

$$\delta^r(T,x) = \left[\ell^r(T,x) - \hat{\ell}^r(T,x)\right] 1\left\{\ell^r(T,x) \geq \hat{\ell}^r(T,x)\right\},$$

and note that $\delta^r(T,x) \leq \ell^r(T,x)$ for all $x \in \mathbb{R}$ and $\delta^r(T,x) \to_{a.s.} 0$ for each $x \in \mathbb{R}$, we have

$$\int_{-\infty}^{\infty} \left|\hat{\ell}^r(T,x) - \ell^r(T,x)\right| dx = 2\int_{-\infty}^{\infty} \delta^r(T,x)\, dx + O(\Delta) \to_{a.s.} 0,$$

due to dominated convergence. The proof is therefore complete. □.

Proof of Corollary 4.4 Following the proof of Theorem 4.2, we define

$$\begin{aligned} R(T,x) &= \hat{L}^r(T,x) - L^r(T,x) \\ &= \sum_{i=1}^n \int_{(i-1)\Delta}^{i\Delta} \left(e^{-ri\Delta} 1\{X_{i\Delta} \leq x\} - e^{-rt} 1\{X_t \leq x\}\right) dt, \end{aligned} \tag{45}$$

and write

$$R(T,x) = R_1(T,x) + R_2(T,x), \tag{46}$$

where

$$R_1(T,x) = \sum_{i=1}^{n} 1\{X_{i\Delta} \le x\} \int_{(i-1)\Delta}^{i\Delta} \left(e^{-ri\Delta} - e^{-rt}\right) dt$$

$$R_2(T,x) = \sum_{i=1}^{n} \int_{(i-1)\Delta}^{i\Delta} e^{-rt} \left(1\{X_{i\Delta} \le x\} - 1\{X_t \le x\}\right) dt.$$

We have, as in the proof of Corollary 4.3,

$$\begin{aligned} |R_1(T,x)| &\le r\Delta \hat{L}(T,x) \\ &\le r\Delta \left[T + \sup_{x\in\mathbb{R}} \left|\hat{L}(T,x) - L(T,x)\right|\right] \\ &\le r\Delta \left[T + 2\omega(\Delta)\left(\sup_{x\in\mathbb{R}} \ell(T,x)\right)\right] \end{aligned} \tag{47}$$

uniformly in $x \in \mathbb{R}$. Moreover, due to (34) and (36), we have

$$|R_2(T,x)| \le \int_0^T 1\{|X_t - x| \le \omega(\Delta)\}\, dt \le 2\omega(\Delta)\left(\sup_{x\in\mathbb{R}} \ell(T,x)\right) \tag{48}$$

uniformly in $x \in \mathbb{R}$. Therefore, it follows from (45), (46), (47) and (48) that

$$\sup_{x\in\mathbb{R}} \left|\hat{L}^r(T,x) - L^r(T,x)\right| \le r\Delta T + 2(1 + r\Delta)\omega(\Delta)\left(\sup_{x\in\mathbb{R}} \ell(T,x)\right), \tag{49}$$

and the proof is complete. □

Proof of Theorem 5.1 Define

$$\tilde{\lambda}_N(T,x) = \frac{1}{N}\sum_{k=1}^{n} \ell_k(T,x),$$

and write

$$\hat{\lambda}_N(T,x) - \lambda(T,x) = \left(\hat{\lambda}_N(T,x) - \tilde{\lambda}_N(T,x)\right) + \left(\tilde{\lambda}_N(T,x) - \lambda(T,x)\right), \tag{50}$$

two terms of which we look at separately below for the proofs of parts (a) and (b).

To prove part (a), note first that we may easily deduce from the ergodic theorem for the strong mixing sequences that

$$\tilde{\lambda}_N(T,x) \to_{a.s.} \lambda(T,x)$$

for all $x \in \mathbb{R}$. Note that $\mathbb{E}\ell_k(T,x) = \lambda(T,x)$ for all $k = 1, 2, \ldots$. Due to (50), the strong pointwise consistency of $\hat{\lambda}_N(T,\cdot)$ in part (a) would therefore follow immediately if we prove

$$\hat{\lambda}_N(T,x) - \tilde{\lambda}_N(T,x) \to 0, \tag{51}$$

for all $x \in \mathbb{R}$ under the given conditions, as $N \to \infty$. To prove (51), we first note that

$$\begin{aligned} \hat{\lambda}_N(T,x) - \tilde{\lambda}_N(T,x) &= \frac{1}{N}\sum_{k=1}^{N}\left[\hat{\ell}_k(T,x) - \ell_k(T,x)\right] \\ &= \frac{1}{N}\sum_{k=1}^{N}\left[\hat{\ell}_k(T,x) - \tilde{\ell}_k(T,x)\right] + \frac{1}{N}\sum_{k=1}^{N}\left[\tilde{\ell}_k(T,x) - \ell_k(T,x)\right], \end{aligned} \tag{52}$$

where $\tilde{\ell}_k(T,x)$ is defined for each X^k similarly as $\tilde{\ell}(T,x)$ in the proof of Theorem 4.1.

For the first term in (52), we use (29) to deduce that

$$\left|\frac{1}{N}\sum_{k=1}^{N}\left[\hat{\ell}_k(T,x) - \tilde{\ell}_k(T,x)\right]\right| \le \frac{\omega(\Delta)}{h^{1+\delta}}\left[c_1 + c_2\frac{1}{N}\sum_{k=1}^{N}\left(\sup_{x\in\mathbb{R}}\ell_k(T,x)\right)\right] \to_{a.s.} 0, \tag{53}$$

if $\omega(\Delta)/h^{1+\delta} \to 0$. Note in particular that the constants c_1 and c_2 do not depend upon k. To consider the second term in (52), note that (31) in the proof of Theorem 4.1 implies

$$\tilde{\ell}_k(T,x) - \ell_k(T,x) = \int_{-\infty}^{\infty} K(s)[\ell_k(T,x+hs) - \ell_k(T,x)]\,ds,$$

which in turn yields,

$$\frac{1}{N}\sum_{k=1}^{N}\left[\tilde{\ell}_k(T,x) - \ell_k(T,x)\right] = \int_{-\infty}^{\infty} ds K(s)\frac{1}{N}\sum_{k=1}^{N}[\ell_k(T,x+hs) - \ell_k(T,x)]. \tag{54}$$

Now we fix s in (54). By the usual arguments to establish the uniform law of large numbers and the law of large numbers for a strong mixing sequence, we may easily derive that

$$\frac{1}{N}\sum_{k=1}^{N}[\ell_k(T,x+hs) - \ell_k(T,x)] \to_{a.s.} \lambda(T,x+hs) - \lambda(T,x) \tag{55}$$

holds uniformly in h locally, i.e., over any compact set including the origin. For the uniform low of large numbers (55), we may just consider the supremum and infimum of $\ell_k(T,x+hs)$ taken over a neighborhood of any given h, and note that $\lambda(T,x+hs)$ is continuous with respect to h. However, we have

$$\lambda(T,x+hs) - \lambda(T,x) \to 0$$

as $h \to 0$, and therefore, if we set $h \to 0$ as $N \to \infty$, then we have

$$\frac{1}{N}\sum_{k=1}^{N}[\ell_k(T,x+hs) - \ell_k(T,x)] \to_{a.s.} 0 \tag{56}$$

as $N \to \infty$ for each s. Consequently, we may deduce from (54), (56) and dominated convergence that

$$\frac{1}{N}\sum_{k=1}^{N}\left[\tilde{\ell}_k(T,x) - \ell_k(T,x)\right] \to_{a.s.} 0 \tag{57}$$

as $N \to \infty$. Now, (51) follows immediately from (52), (53) and (57), which establishes the strong pointwise consistency of $\hat{\lambda}_N(T, \cdot)$. The L^1-consistency of $\hat{\lambda}_N(T, \cdot)$ can be established exactly as in the proof of Theorem 4.1, following the proof of Scheffé's theorem. The proof for part (a) is therefore complete.

For part (b), we first show that

$$\sqrt{N}\left(\hat{\lambda}_N(T, x) - \tilde{\lambda}_N(T, x)\right) \to_{a.s.} 0 \tag{58}$$

for all $x \in \mathbb{R}$ as $N \to \infty$, under the given conditions. To show (58), we note similarly as in (53) that

$$\left|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\left[\hat{\ell}_k(T, x) - \tilde{\ell}_k(T, x)\right]\right| \leq \sqrt{N}\,\frac{\omega(\Delta)}{h^{1+\delta}}\left[c_1 + c_2\frac{1}{N}\sum_{k=1}^{N}\left(\sup_{x\in\mathbb{R}}\ell_k(T, x)\right)\right] \to_{a.s.} 0, \tag{59}$$

if $\omega(\Delta)/h^{1+\delta} = o(N^{-1/2})$. Moreover, we have

$$|\ell_k(T, x + hs) - \ell_k(T, x)| \leq C_k|hs|^{1/2-\delta}$$

for some sequence of random variables C_k, due to the Hölder continuity of $\ell_k(T, \cdot)$. Therefore, using the result in (54), we may easily deduce that

$$\begin{aligned}\left|\frac{1}{\sqrt{N}}\sum_{k=1}^{N}\left[\tilde{\ell}_k(T, x) - \ell_k(T, x)\right]\right| &\leq \int_{-\infty}^{\infty} ds\, K(s)\,\frac{1}{\sqrt{N}}\sum_{k=1}^{N}|\ell_k(T, x + hs) - \ell_k(T, x)| \\ &\leq \left(h^{1/2-\delta}N^{1/2}\right)\frac{1}{N}\sum_{k=1}^{N}C_k\int_{-\infty}^{\infty}|s|^{1/2-\delta}K(s)\,ds \\ &\to_{a.s.} 0 \end{aligned} \tag{60}$$

under the given condition $h^{1-\delta} = o(N^{-1})$ and $\int_{-\infty}^{\infty}|s|^{1/2}K(s)\,ds < \infty$. Now, (58) follows follows immediately from (59) and (60), due to (52).

However, we have from the central limit theorem for the strong mixing sequence

$$\sqrt{N}\left(\tilde{\lambda}_N(T, x) - \lambda(T, x)\right) = \frac{1}{\sqrt{N}}\sum_{k=1}^{N}[\ell_k(T, x) - \lambda(T, x)] \tag{61}$$

converges in distribution to normal law with variance

$$\sigma_S^2(T, x) = \lim_{N\to\infty}\frac{1}{N}\mathbb{E}\left(\sum_{k=1}^{N}[\ell_k(T, x) - \lambda(T, x)]\right)^2.$$

See, e.g., Hall and Heyde (1980, Corollary 5.1, p132). Bosq (1998, Theorem 1.7, p36) also provides a similar result with a slightly stronger condition. The stated result in part (b) therefore follows readily from (61) and (58), and the proof is complete. □

Proof of Theorem 5.2 Let

$$L_k(T,x) = \int_0^T 1\{X_t^k \le x\}\,dt$$

and

$$\Lambda_N(T,x) = \frac{1}{N}\sum_{k=1}^N L_k(T,x).$$

To prove part (a), we write

$$\hat{\Lambda}_N(T,x) - \Lambda(T,x) = \left(\hat{\Lambda}_N(T,x) - \Lambda_N(T,x)\right) + (\Lambda_N(T,x) - \Lambda(T,x)), \qquad (62)$$

each term of which we consider below. For the second term in (62), we may deduce from the ergodic theorem for the strong mixing sequences that

$$\Lambda_N(T,x) \to_{a.s.} \Lambda(T,x)$$

for each $x \in \mathbb{R}$ as $N \to \infty$. Note that $\mathbb{E}L_k(T,x) = \Lambda(T,x)$ for all $k = 1, 2, \ldots$. We may then readily establish that

$$\sup_{x\in\mathbb{R}} |\, \Lambda_N(T,x) - \Lambda(T,x)\,| \to_{a.s.} 0 \qquad (63)$$

as $N \to \infty$, as in the proof of the classical Glivenko-Cantelli theorem [see, e.g., Durrett (1991, p56)]. For the first term in (62), we write

$$\hat{\Lambda}_N(T,x) - \Lambda_N(T,x) = \frac{1}{N}\sum_{k=1}^N \left(\hat{L}_k(T,x) - L_k(T,x)\right).$$

However, we have

$$\sup_{x\in\mathbb{R}} \left|\hat{L}_k(T,x) - L_k(T,x)\right| \le 2\omega(\Delta)\left(\sup_{x\in\mathbb{R}} \ell_k(T,x)\right)$$

as shown in (37), and therefore,

$$\sup_{x\in\mathbb{R}} \left|\hat{\Lambda}_N(T,x) - \Lambda_N(T,x)\right| \le 2\omega(\Delta)\frac{1}{N}\sum_{k=1}^N \left(\sup_{x\in\mathbb{R}} \ell_k(T,x)\right). \qquad (64)$$

The result stated in part (a) now follows directly from (62), (63) and (64).

For the proof of part (b), we write

$$\begin{aligned} \hat{U}_N(T,x) &= \sqrt{N}\left(\hat{\Lambda}_N(T,x) - \Lambda(T,x)\right) \\ &= \sqrt{N}\left(\hat{\Lambda}_N(T,x) - \Lambda_N(T,x)\right) + \sqrt{N}\left(\Lambda(T,x) - \Lambda(T,x)\right). \end{aligned} \qquad (65)$$

We may easily deduce from (64) that

$$\sqrt{N}\sup_{x\in\mathbb{R}} \left|\hat{\Lambda}_N(T,x) - \Lambda_N(T,x)\right| \le 2\sqrt{N}\omega(\Delta)\frac{1}{N}\sum_{k=1}^N \left(\sup_{x\in\mathbb{R}} \ell_k(T,x)\right) \to_{a.s.} 0 \qquad (66)$$

under the given condition. It now suffices to show that

$$U_N(T,\cdot) \to_d U(T,\cdot) \tag{67}$$

as $N \to \infty$, where

$$U_N(T,x) = \sqrt{N}\,(\Lambda_N(T,x) - \Lambda(T,x)) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \Big(L_k(T,x) - \Lambda(T,x) \Big),$$

due to (65) and (66).

Our proof of (67) uses the approaches by Pollard (1990) and Andrews and Pollard (1994). For each $x \in \mathbb{R}$, define a functional $f : C[0,T] \to \mathbb{R}$ by

$$f(\,\cdot\,) = \int_0^T 1\{\,\cdot\, \leq x\}\, dt,$$

so that we have in particular $f(X^k) = L_k(T,x)$, and let $\mathcal{F}$ be the set of all such functionals defined for all $x \in \mathbb{R}$. For brevity, we will identify x with its associated functional in what follows. We introduce a pseudometric ρ on $\mathcal{F}$ defined by

$$\rho^2(x,y) = \mathbb{E}\Big| L_k(T,x) - L_k(T,y) \Big|^2, \tag{68}$$

and view $U_N(T,\cdot)$ and $U(T,\cdot)$ as stochastic processes indexed by $(\mathcal{F},\rho)$. Due to Pollard (1990, Theorem 10.2), it now suffices to show that

(a) the finite dimensional distributions of $U_N(T,\cdot)$ converge weakly to those of $U(T,\cdot)$,

(b) the pseudometric space $(\mathcal{F},\rho)$ is totally bounded, and $U_N(T,\cdot)$ is stochastically equicontinuous,

to derive (67).

The weak convergence of the finite dimensional distributions $U_N(T,\cdot)$ to those of $U(T,\cdot)$ in (a) follows directly from the central limit theorem for the strong mixing sequences. See, e.g., Bosq (1998, Theorem 1.7, p36) or Hall and Heyde (1980, Corollary 5.1, p132). Note in particular that the latter, if applied with $\delta = \infty$, actually yields (a) without any condition on the mixing coefficients except for absolute summability. To establish (b), we will verify the mixing and bracketing conditions in Andrews and Pollard (1994, Theorem 2.2). To obtain the bracketing numbers $N(\varepsilon,\mathcal{F})$ for $\mathcal{F}$, we let $\varepsilon > 0$ be given arbitrarily. Furthermore, we define

$$x_i = \inf_{x \in \mathbb{R}} \left\{ \Lambda(T,x) \geq \frac{i\varepsilon^2}{T} \right\}$$

for $i = 1, \ldots, [T/\varepsilon^2]$, and denote by

$$I_i = [x_{i-1}, x_i]$$

for $i = 1, \ldots, [T/\varepsilon^2] + 1$ with the convention that $x_0 = -\infty$ and $x_{[T/\varepsilon^2]+1} = \infty$. Then it follows that

$$\begin{aligned}\mathbb{E} \sup_{x,y \in I_i} \left| L_k(T,x) - L_k(T,y) \right|^2 &= \mathbb{E} \sup_{x,y \in I_i} \left(\int_0^T 1\{x < X_t^k \le y\}\, dt \right)^2 \\ &\le \mathbb{E} \left(\int_0^T 1\{x_{i-1} < X_t^k \le x_i\}\, dt \right)^2 \\ &\le T\mathbb{E} \int_0^T 1\{x_{i-1} < X_t^k \le x_i\}\, dt = \varepsilon^2, \end{aligned} \tag{69}$$

where we assume $x < y$ without loss of generality. Consequently,

$$N(\varepsilon, \mathcal{F}) = [T/\varepsilon^2] + 1$$

for any $\varepsilon > 0$ given.

To employ the result by Andrews and Pollard (1994, Theorem 2.2), we need to show that

$$\sum_{k=1}^n k^{a-2} \alpha(k)^{b/(a+b)} < \infty \quad \text{and} \quad \int_0^1 x^{-b/(2+b)} N(x, \mathcal{F})^{1/a} dx < \infty \tag{70}$$

for some even integers $a \ge 2$ and $b > 0$. Let $\alpha(k) = k^{-c}$ and $N(x, \mathcal{F}) = x^{-2}$. Then the conditions in (70) are satisfied if and only if

$$(a-2) - \frac{cb}{a+b} < -1, \quad -\frac{2}{a} - \frac{b}{2+b} > -1$$

which hold if and only if

$$\frac{a(a-1)}{c-(a-1)} < b < a-2$$

Therefore, the required a and b exist if and only if

$$c > \frac{2(a-1)^2}{a-2}$$

In particular, we need $c > 9$ as assumed, if we set $a = 4$. This proves (67), and the proof for part (b) is therefore complete. □

Proof of Corollary 5.3 The proof is a straightforward multivariate extension of that of Theorem 5.2. The details are therefore omitted. □

Proof of Corollary 5.4 The proofs of the stated results are parallel to those of Theorems 5.1 and 5.2, and Corollary 5.3. We first derive the results for $\hat{\lambda}^r(T,x)$ corresponding to those in part (a) of Theorem 5.1. We have (53)

$$\frac{1}{N} \sum_{k=1}^N \left[\hat{\ell}_k^r(T,x) - \tilde{\ell}_k^r(T,x) \right] \to_{a.s.} 0 \tag{71}$$

as $N \to \infty$, which follows from (43). Moreover, it follows from Lemmas A2 and A3 that $\ell_k^r(T,\cdot)$ is hölder continuous of order $1/2-\delta$ and $\lambda^r(T,\cdot)$ is continuous, and therefore, we have

$$\frac{1}{N}\sum_{k=1}^N \left[\tilde{\ell}_k^r(T,x) - \ell_k^r(T,x)\right] \to_{a.s.} 0 \tag{72}$$

as $N \to \infty$. The results in (71) and (72) correspond respectively to (53) and (57). The strong pointwise consistency of $\hat{\lambda}_N^r(T,\cdot)$ is therefore easily established as in the proof of Theorem 5.1. For the L^1-consistency of $\hat{\lambda}_N^r(T,\cdot)$ can be obtained exactly as in the proof of Corollary 4.3, given the strong pointwise consistency of $\hat{\lambda}_N^r(T,\cdot)$.

For the results of $\hat{\lambda}^r(T,x)$ comparable to those in part (b) of Theorem 5.1, we show

$$\frac{1}{\sqrt{N}}\sum_{k=1}^N \left[\hat{\ell}_k^r(T,x) - \tilde{\ell}_k^r(T,x)\right] \to_{a.s.} 0 \tag{73}$$

from (43), and

$$\frac{1}{\sqrt{N}}\sum_{k=1}^N \left[\tilde{\ell}_k^r(T,x) - \ell_k^r(T,x)\right] \to_{a.s.} 0 \tag{74}$$

using the Hölder continuity of $\ell_k^r(T,\cdot)$ in Lemma A3, and apply the central limit theorem for the strong mixing sequence

$$\sqrt{N}\left(\tilde{\lambda}_N^r(T,x) - \lambda^r(T,x)\right) = \frac{1}{\sqrt{N}}\sum_{k=1}^N \left[\ell_k^r(T,x) - \lambda^r(T,x)\right]. \tag{75}$$

The actual derivations required in (73), (74) and (75) are essentially identical to those in (59), (60) and (61) in the proof of Theorem 5.1. Note that we have under the given conditon $\mathbb{E}|\ell_k^r(T,x)|^{2+\delta} \le \mathbb{E}|\ell_k(T,x)|^{2+\delta} < \infty$.

To derive the results of $\hat{\Lambda}^r(T,x)$ that are similar to those in part (a) of Theorem 5.2, we employ the ergodic theorem for the strong mixing sequences and the proof of the Glivenko-Cantelli theorem to establish

$$\sup_{x\in\mathbb{R}} |\Lambda_N^r(T,x) - \Lambda^r(T,x)| \to_{a.s.} 0, \tag{76}$$

precisely as in (63). Moreover, we deduce

$$\sup_{x\in\mathbb{R}} \left|\hat{\Lambda}_N^r(T,x) - \Lambda_N^r(T,x)\right| \to_{a.s.} 0 \tag{77}$$

as in (64), using the result in (49). The required results of $\hat{\Lambda}^r(T,x)$ now follow immediately from (76) and (77), as in the proof of Theorem 5.2.

For the results of $\hat{\Lambda}^r(T,x)$ corresponding to those in part (b) of Theorem 5.2, we deduce from (49) that

$$\sqrt{N}\sup_{x\in\mathbb{R}} \left|\hat{\Lambda}_N^r(T,x) - \Lambda_N^r(T,x)\right| \to_{a.s.} 0, \tag{78}$$

and show that

$$U_N^r(T,\cdot) \to_d U^r(T,x), \tag{79}$$

where $U_N^r(T,\cdot)$ is defined similarly as $U_N(T,\cdot)$ from $\Lambda_N^r(T,\cdot)$ and $\Lambda^r(T,x)$. The details for the proofs of (78) and (79) are virtually identical to those of (66) and (67) in the proof of Theorem 5.2. For (79), we need some obvious modifications in the proof of (67). The required modification, however, is pretty straightforward. In fact, it is only necessary to redefine

$$\rho^2(x,y) = \mathbb{E}\,|L_k^r(T,x) - L_k^r(T,y)|$$

in (68), and show that

$$\mathbb{E} \sup_{x,y\in I_i} |L_k^r(T,x) - L_k^r(T,y)| \leq \varepsilon^2, \tag{80}$$

where $I_i = [x_{i-1}, x_i]$ with x_i's given by

$$x_i = \inf_{x\in\mathbb{R}} \left\{ \Lambda^r(T,x) \geq \frac{i\varepsilon^2}{T^r} \right\} \quad \text{and} \quad T^r = \int_0^T e^{-rt} dt,$$

for $i = 1, \ldots, [T^r/\varepsilon^2]$. The derivation of (80) is entirely analogous to that of (69) in the proof of Theorem 5.2. Finally, the result for $(\hat{\Lambda}_N^{r,X}, \hat{\Lambda}_N^{r,Y})$ is a straightforward multivariate extension of the univariate case, just as for the case of Corollary 5.3. The proof is therefore complete. □

Proof of Theorem 5.5

The proofs of the stated results are rather involved, and therefore, will be done in several steps. We will first present some lemmas that will be used repeatedly in the proofs. The proofs of the stated results will then be given subsequently.

Useful Lemmas and Their Proofs

Lemma 5.5.1 For all y such that $|x-y| \leq \omega(\underline{\Delta})$, we have

$$\underline{\Delta}^{1+\delta} \int_{-\infty}^{\infty} dx \left| \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) p(([mt]-k)\underline{\Delta}, y, z) - \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) p(([mt]-k)\underline{\Delta}, x, z) \right| = o(\omega(\underline{\Delta})) \text{ a.s.}$$

uniformly in t, y, z, as $\Delta \to 0$.

Proof of Lemma 5.5.1 The stated result follows directly from

$$\begin{aligned}
&\underline{\Delta}^{1+\delta}\int_{-\infty}^{\infty} dx \left|\sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x)p(([mt]-k)\underline{\Delta}, y, z) - \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x)p(([mt]-k)\underline{\Delta}, x, z)\right| \\
&\qquad \le \omega(\underline{\Delta})\underline{\Delta}^{1+\delta}c_1 \int_{-\infty}^{\infty} dx \sum_{k=0}^{[mt]-1} \frac{p(k\underline{\Delta}, x_0, x)}{([mt]-k)\underline{\Delta}} \\
&\qquad \le \omega(\underline{\Delta})\underline{\Delta}^{\delta}c_1 \sum_{k=1}^{[mt]} \frac{1}{k} \\
&\qquad \le \omega(\underline{\Delta})\underline{\Delta}^{\delta}c_1 \left(1 + \int_1^{[mt]} \frac{1}{s} ds\right) \\
&\qquad \le \omega(\underline{\Delta})\underline{\Delta}^{\delta}c_1 \left(1 + \log\frac{t}{\underline{\Delta}}\right).
\end{aligned}$$

In particular, the first inequality is due to the first-order Taylor expansion and Assumption 5.5(a), and the second inequality follows from the fact that $\int_{-\infty}^{\infty} p(k\underline{\Delta}, x_0, x)\,dx = 1$ for all k. □

Lemma 5.5.2 We have

$$\begin{aligned}
&\left|p(\underline{\Delta}, x + hw, y + \underline{h}z) - p(\underline{\Delta}, x, y) - \left(hw\frac{\partial}{\partial x} + \underline{h}z\frac{\partial}{\partial y}\right)p(\underline{\Delta}, x, y)\right| \\
&\qquad \le c\frac{h^2w^2 + h\underline{h}|w||z| + \underline{h}^2z^2}{\underline{\Delta}^{3/2}}
\end{aligned}$$

for some constant $c > 0$, uniformly in w, x, y, z.

Proof of Lemma 5.5.3 From the second-order Taylor expansion, we have

$$\begin{aligned}
p(\underline{\Delta}, x + hw, y + \underline{h}z) \;=&\; p(\underline{\Delta}, x, y) + \left(hw\frac{\partial}{\partial x} + \underline{h}z\frac{\partial}{\partial y}\right)p(\underline{\Delta}, x, y) \\
&+ \left(\frac{1}{2}h^2w^2\frac{\partial^2}{\partial x^2} + h\underline{h}wz\frac{\partial^2}{\partial x\partial y} + \frac{1}{2}\underline{h}^2z^2\frac{\partial^2}{\partial y^2}\right)p(\underline{\Delta}, x^*, y^*)
\end{aligned}$$

for some x^* and y^*. Therefore, the stated result follows immediately from Assumption 5.5(a). □

Lemma 5.5.3 For any transformation F on $\mathbb{R}$ such that it is infinitely differentiable with bounded and absolutely integrable derivatives and $\int_{-\infty}^{\infty} |s|^{1/2-\delta}|F(s)|\,ds < \infty$, we have

$$\frac{\Delta}{h}\sum_{i=1}^{n\kappa_x} F\left(\frac{X_{i\Delta} - x}{h}\right) = M\int_{-\infty}^{\infty} F(s)\,ds + O\left(\frac{\omega(\Delta)}{h^{1+\delta}}M\right) + O(h^{1/2-\delta}M)$$

uniformly in x, as $\Delta \to 0$.

Proof of Lemma 5.5.3 We may deduce exactly as in (29) that

$$\frac{\Delta}{h}\sum_{i=1}^{n\kappa_x} F\left(\frac{X_{i\Delta}-x}{h}\right) = \frac{1}{h}\int_0^{\kappa_x T} F\left(\frac{X_t - x}{h}\right) dt + O\left(\frac{\omega(\Delta)}{h^{1+\delta}} M\right)$$

for any function F that is infinitely differentiable and has bounded and absolutely integrable derivatives. Moreover, we have

$$\frac{1}{h}\int_0^{\kappa_x T} F\left(\frac{X_t - x}{h}\right) dt = \frac{1}{h}\int_{-\infty}^{\infty} F\left(\frac{s-x}{h}\right) \ell(\kappa_x T, s)\, ds = \int_{-\infty}^{\infty} F(s)\ell(\kappa_x T, x+hs)\, ds$$

by the successive applications of occupation times formula and change-of-variables. The stated result now follows from

$$\begin{aligned}\int_{-\infty}^{\infty} & F(s)\ell(\kappa_x T, x+hs)\, ds \\ &= \ell(\kappa_x T, x)\int_{-\infty}^{\infty} F(s)\, ds + \int_{-\infty}^{\infty} F(s)\left[\ell(\kappa_x T, x+hs) - \ell(\kappa_x T, x)\right] ds\end{aligned}$$

and

$$\begin{aligned}\int_{-\infty}^{\infty} & F(s)\left[\ell(\kappa_x T, x+hs) - \ell(\kappa_x T, x)\right] ds \\ &\leq cMh^{1/2-\delta}\int_{-\infty}^{\infty} |s|^{1/2-\delta}|F(s)|\, ds = O(h^{1/2-\delta}M)\end{aligned}$$

for some constant $c > 0$. □

Proofs of the Main Results

The proofs of the main results will be presented in four steps, which will be given subsequently below. In our subsequent proofs, δ denotes an arbitrary small nonnegative number, which may vary from line to line. Note that we have

$$\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta} \to \infty \quad (81)$$

$$\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2} \frac{\Delta^{1/2-\delta} + h^{3/2-\delta}}{\underline{\Delta}^{1/2}} \to 0 \quad (82)$$

$$\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2} \frac{h^{2-\delta}}{\underline{\Delta}} \to 0 \quad (83)$$

$$\frac{h}{\underline{\Delta}^{1/2}} \to 0 \quad (84)$$

under the given conditions for $h, \underline{h}, \Delta, \underline{\Delta}$ and M.

First Step In the first step, we obtain a decomposition of $\hat{p}_\Delta(\underline{\Delta}, x, y)$ that will be used throughout the proof. To obtain the desired decomposition, note that $\hat{p}_\Delta(\underline{\Delta}, x, y)$ is the standard Nadaraya-Watson kernel estimator, based on the sample of size $n\kappa_x$, of the conditional mean function

$$\begin{aligned}\mathbb{E}\left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)\middle| X_{i\Delta}=x\right] &= \frac{1}{\underline{h}}\int_{-\infty}^{\infty}K\left(\frac{z-y}{\underline{h}}\right)p(\underline{\Delta},x,z)\,dz\\ &= \int_{-\infty}^{\infty}K(z)p(\underline{\Delta},x,y+\underline{h}z)\,dz.\end{aligned}$$

Therefore, we may write

$$\begin{aligned}\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right) &= \int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\\ &\quad+\left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right],\end{aligned} \tag{85}$$

and decompose $(1/\underline{h})K((X_{\underline{\Delta}+i\Delta}-y)/\underline{h})$ into the conditional mean and the martingale difference error. Consequently, we may define

$$\hat{p}_\Delta(\underline{\Delta},x,y)=\tilde{p}_\Delta(\underline{\Delta},x,y)+\tilde{q}_\Delta(\underline{\Delta},x,y), \tag{86}$$

where

$$\tilde{p}_\Delta(\underline{\Delta},x,y)=\frac{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}K\left(\frac{X_{i\Delta}-x}{h}\right)\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz}{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}K\left(\frac{X_{i\Delta}-x}{h}\right)}$$

$$\tilde{q}_\Delta(\underline{\Delta},x,y)=\frac{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}K\left(\frac{X_{i\Delta}-x}{h}\right)\left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right]}{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}K\left(\frac{X_{i\Delta}-x}{h}\right)}$$

correspondingly as the decomposition in (85).

We now let $m=1/\underline{\Delta}$ and notice that

$$\begin{aligned}&\left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2}\left[\hat{p}_\Delta(t,x_0,z)-p(t,x_0,z)\right]\\ &=\left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2}\sum_{k=0}^{[mt]}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\\ &\quad p(k\underline{\Delta},x_0,x)\left[\hat{p}_\Delta(\underline{\Delta},x,y)-p(\underline{\Delta},x,y)\right]p(([mt]-k)\underline{\Delta},y,z)+o_p(1)\end{aligned} \tag{87}$$

uniformly in t, as $\Delta\to 0$. In (87), we use the convention $p(0,\cdot,\cdot)=1$, which will be made throughout the proof. Quite clearly, all the other terms include repeated products of

$\hat{p}_\Delta(\underline{\Delta},\cdot,\cdot) - p(\underline{\Delta},\cdot,\cdot)$, and therefore, are of order small than the leading term given in (87). Therefore, we may write

$$\left(\frac{Mh\underline{h}\Delta^2}{\Delta}\right)^{1/2}[\hat{p}_\Delta(t,x_0,z) - p(t,x_0,z)] = g_\Delta(t,z) + f_\Delta(t,z) + o_p(1), \tag{88}$$

where

$$\begin{aligned} g_\Delta(t,z) &= \left(\frac{Mh\underline{h}\Delta^2}{\Delta}\right)^{1/2}\sum_{k=0}^{[mt]}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy \\ &\quad p(k\underline{\Delta},x_0,x)\,[\tilde{p}_\Delta(\underline{\Delta},x,y) - p(\underline{\Delta},x,y)]\,p(([mt]-k)\underline{\Delta},y,z) \\ f_\Delta(t,z) &= \left(\frac{Mh\underline{h}\Delta^2}{\Delta}\right)^{1/2}\sum_{k=0}^{[mt]}\int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\,p(k\underline{\Delta},x_0,x)\tilde{q}_\Delta(\underline{\Delta},x,y)p(([mt]-k)\underline{\Delta},y,z) \end{aligned}$$

due to (86) and (87). □

Second Step As the second step, we will establish that

$$g_\Delta(t,z) = o_p(1) \tag{89}$$

uniformly in t and z, as $\Delta \to 0$. However, we have from Lemma 5.5.1 that

$$\begin{aligned} |g_\Delta(t,z)| &\le \left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\left(\sup_{x,y\in\mathbb{R}}|\tilde{p}_\Delta(\underline{\Delta},x,y) - p(\underline{\Delta},x,y)|\right) \\ &\quad \int_{-\infty}^{\infty}dx\int_{-\infty}^{\infty}dy\,\underline{\Delta}\sum_{k=0}^{[mt]}p(k\underline{\Delta},x_0,x)p(([mt]-k)\underline{\Delta},y,z)\,1\{|x-y|\le\omega(\underline{\Delta})\} \\ &\le 2\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\omega(\underline{\Delta})\left(\sup_{x,y\in\mathbb{R}}|\tilde{p}_\Delta(\underline{\Delta},x,y) - p(\underline{\Delta},x,y)|\right) \\ &\quad \left[\int_{-\infty}^{\infty}dx\left(\underline{\Delta}\sum_{k=0}^{[mt]}p(k\underline{\Delta},x_0,x)p(([mt]-k)\underline{\Delta},x,z)\right)\right. \\ &\quad \left. + \int_{-\infty}^{\infty}dx\left(\underline{\Delta}\sum_{k=0}^{[mt]}p(k\underline{\Delta},x_0,x)|p(([mt]-k)\underline{\Delta},x,z) - p(([mt]-k)\underline{\Delta},y,z)|\right)\right] \\ &\le 2\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\omega(\underline{\Delta})\left(\sup_{x,y\in\mathbb{R}}|\tilde{p}_\Delta(\underline{\Delta},x,y) - p(\underline{\Delta},x,y)|\right) \\ &\quad \left[\left(\int_{-\infty}^{\infty}dx\int_0^t ds\,p(u,x_0,x)p(t-s,x,z)\right) + o(1)\right] \text{ a.s.} \end{aligned}$$

uniformly in t and z.

To establish (89), it therefore suffices to show that

$$\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\omega(\underline{\Delta})\sup_{x,y\in\mathbb{R}}|\tilde{p}_\Delta(\underline{\Delta},x,y)-p(\underline{\Delta},x,y)|\to_{a.s.}0 \tag{90}$$

as $\Delta\to 0$. To show (90), we let

$$w=\frac{X_{i\Delta}-x}{h}$$

so that $X_{i\Delta}=x+hw$. Then it follows from Lemma 5.5.2 that

$$\tilde{p}_\Delta(\underline{\Delta},x,y)-p(\underline{\Delta},x,y)-h\left(\frac{\partial}{\partial x}p(\underline{\Delta},x,y)\right)\left[O\left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)+O(h^{1/2-\delta})\right]=O\left(\frac{h^2}{\underline{\Delta}^{3/2}}\right)\text{ a.s.} \tag{91}$$

uniformly in x and y, using in particular the facts that $\int_{-\infty}^{\infty}sK(s)\,ds=0$, $\int_{-\infty}^{\infty}s^2K(s)\,ds<\infty$, $\underline{h}=O(h)$,

$$\frac{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}\left(\frac{X_{i\Delta}-x}{h}\right)K\left(\frac{X_{i\Delta}-x}{h}\right)}{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}K\left(\frac{X_{i\Delta}-x}{h}\right)}=O\left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)+O(h^{1/2-\delta})\text{ a.s.}$$

and

$$\frac{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}\left|\frac{X_{i\Delta}-x}{h}\right|^kK\left(\frac{X_{i\Delta}-x}{h}\right)}{\dfrac{\Delta}{h}\displaystyle\sum_{i=1}^{n\kappa_x}K\left(\frac{X_{i\Delta}-x}{h}\right)}=\int_{-\infty}^{\infty}|s|^kK(s)\,ds+O\left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)+O(h^{1/2-\delta})\text{ a.s.}$$

for $k=1,2$, which are due to Lemma 5.5.4. Consequently, we have from (91)

$$\sup_{x,y\in\mathbb{R}}|\tilde{p}_\Delta(\underline{\Delta},x,y)-p(\underline{\Delta},x,y)|=O\left(\frac{h^{-\delta}\omega(\Delta)+h^{3/2-\delta}}{\underline{\Delta}}\right)+O\left(\frac{h^2}{\underline{\Delta}^{3/2}}\right)\text{ a.s.}$$

uniformly in x and y, due in particular to Assumption 5.5(a). We therefore have

$$\begin{aligned}&\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\omega(\underline{\Delta})\left(\sup_{x,y\in\mathbb{R}}|\tilde{p}_\Delta(\underline{\Delta},x,y)-p(\underline{\Delta},x,y)|\right)\\&=O\left(\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\frac{\Delta^{1/2-\delta}+h^{3/2-\delta}}{\underline{\Delta}^{1/2}}\right)+O\left(\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\frac{h^{2-\delta}}{\underline{\Delta}^{1/2}}\right)=o(1)\text{ a.s.}\end{aligned}$$

uniformly in x and y, due to conditions (82) and (83). This is what was to be shown to prove (90). □

Third Step We now investigate the asymptotics for $f_\Delta(t,z)$, for which we introduce a continuous martingale $Q^\Delta(x,y)$ on $[0,\kappa_x T]$ defined for each $x,y\in\mathbb{R}$ by

$$Q_t^\Delta(x,y)=\sqrt{\frac{\underline{h}\Delta}{Mh}}\left(\sum_{i=1}^{k}K\left(\frac{X_{i\Delta}-x}{h}\right)\left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right]\right.$$
$$\left.+K\left(\frac{X_{k\Delta}-x}{h}\right)\left[\frac{1}{\underline{h}}K\left(\frac{X_{t(\underline{\Delta}/\Delta)+k\Delta}-y}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(t(\underline{\Delta}/\Delta),X_{k\Delta},y+\underline{h}z)\,dz\right]\right)$$

for $k\Delta<t\le(k+1)\Delta$, $k=0,\dots,n\kappa_x-1$. It can be readily deduced that

$$\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2}\tilde{q}_\Delta(\underline{\Delta},x,y)=Q^\Delta_{\kappa_x T}(x,y)\left(1+O\left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)+O(h^{1/2-\delta})\right)\ \text{a.s.}\tag{92}$$

uniformly in x and y, due to Lemma 5.5.4. The limiting distribution of $\tilde{q}_\Delta(\underline{\Delta},x,y)$ is therefore given by the continuous martingale $Q^\Delta(x,y)$.

Let $[\cdot]$ and $[\cdot,\cdot]$ respectively be the quadratic variation and covariation of continuous martingales. Subsequently, we will obtain $[Q^\Delta(x,y)]$ for $(x,y)\in\mathbb{R}^2$ and $[Q^\Delta(x,y),Q^\Delta(u,v)]$ for $(x,y),(u,v)\in\mathbb{R}^2$ such that $(x,y)\neq(u,v)$ to find the limiting distributions of the family of continuous martingales $Q^\Delta(x,y)$, and consequently, those of their functionals. To derive $[Q^\Delta(x,y)]$, we first note

$$\underline{h}\mathbb{E}\left[\left(\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right)^2\middle|\,X_{i\Delta}=x\right]$$
$$=\int_{-\infty}^{\infty}K^2(z)p(\underline{\Delta},x,y+\underline{h}z)\,dz-\underline{h}\left[\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},x,y+\underline{h}z)\,dz\right]^2$$

to deduce that

$$[Q^\Delta(x,y)]_{\kappa_x T}=\frac{\Delta}{Mh}\sum_{i=1}^{n\kappa_x}K^2\left(\frac{X_{i\Delta}-x}{h}\right)$$
$$\left[\int_{-\infty}^{\infty}K^2(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz-\underline{h}\left(\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right)^2\right].\tag{93}$$

However, we have exactly as in (91)

$$\frac{\Delta}{Mh}\sum_{i=1}^{n\kappa_x}K^2\left(\frac{X_{i\Delta}-x}{h}\right)\int_{-\infty}^{\infty}K^2(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz$$
$$=p(\underline{\Delta},x,y)\left(\int_{-\infty}^{\infty}K^2(s)\,ds\right)^2\left(1+O\left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)+O(h^{1/2-\delta})\right)$$
$$+O\left(\frac{h^{-\delta}\omega(\Delta)+h^{3/2-\delta}}{\underline{\Delta}}\right)+O\left(\frac{h^2}{\underline{\Delta}^{3/2}}\right)\ \text{a.s.}$$

uniformly in x and y, due to Lemmas 5.5.3 and 5.5.4 and the required conditions $\int_{-\infty}^{\infty} sK^2(s)\,ds = 0$ and $\int_{-\infty}^{\infty} s^2K^2(s)\,ds < \infty$. Moreover, we may similarly deduce that

$$\begin{aligned}
&\frac{\Delta}{Mh}\sum_{i=1}^{n\kappa_x} K^2\left(\frac{X_{i\Delta}-x}{h}\right)\underline{h}\left(\int_{-\infty}^{\infty} K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right)^2\\
&\quad= \underline{h}p^2(\underline{\Delta},x,y)\left(1+O\left(\frac{\omega(\Delta)}{h^{1+\delta}}\right)+O(h^{1/2-\delta})\right)\\
&\qquad+O\left(\underline{h}p(\underline{\Delta},x,y)\frac{h^{-\delta}\omega(\Delta)+h^{3/2-\delta}}{\underline{\Delta}}\right)+O\left(\underline{h}p(\underline{\Delta},x,y)\frac{h^2}{\underline{\Delta}^{3/2}}\right)\\
&\quad= o(p(\underline{\Delta},x,y))+o\left(\frac{h^{-\delta}\omega(\Delta)+h^{3/2-\delta}}{\underline{\Delta}}\right)+o\left(\frac{h^2}{\underline{\Delta}^{3/2}}\right)\ \text{a.s.}
\end{aligned}$$

uniformly in x and y, if we assume $\int_{-\infty}^{\infty} sK^2(s)\,ds = 0$ and $\int_{-\infty}^{\infty} s^4K^2(s)\,ds < \infty$. Note from Assumption 5.5(a) and condition (84) on $\underline{h}$ and $\underline{\Delta}$ that $p(\underline{\Delta},x,y) \le c_0\underline{\Delta}^{-1/2}$ and $\underline{h}/\underline{\Delta}^{1/2} = o(1)$, which implies in particular that

$$\underline{h}p(\underline{\Delta},x,y) = o(1)\ \text{a.s.}$$

uniformly in x and y. Consequently, we may conclude from (93) that

$$[Q^{\Delta}(x,y)]_{\kappa_x T} = \kappa_2^2\, p(\underline{\Delta},x,y)[1+o(1)]\ \text{a.s.} \tag{94}$$

uniformly in x and y, under our conditions.

Next we consider $[Q^{\Delta}(x,y),Q^{\Delta}(u,v)]$. First, we look at the case $x = u$ and $y \ne v$. We have

$$\begin{aligned}
&\underline{h}\mathbb{E}\left[\left(\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\right)\right.\\
&\qquad\left.\left(\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-v}{\underline{h}}\right)-\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},v+\underline{h}z)\,dz\right)\right| X_{i\Delta}=x\Bigg]\\
&\quad=\int_{-\infty}^{\infty}K(z)K\left(z+\frac{y-v}{\underline{h}}\right)p(\underline{\Delta},x,y+\underline{h}z)\,dz\\
&\qquad-\underline{h}\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},x,y+\underline{h}z)\,dz\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},x,v+\underline{h}z)\,dz.
\end{aligned}\tag{95}$$

However, it can be shown that for all v outside a neighborhood of y

$$\frac{\Delta}{Mh}\sum_{i=1}^{k}K^2\left(\frac{X_{i\Delta}-x}{h}\right)\int_{-\infty}^{\infty}K(z)K\left(z+\frac{y-v}{\underline{h}}\right)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz = o(p(\underline{\Delta},x,y))\ \text{a.s.}$$

uniformly in $1 \le k \le n\kappa_x$, and in x and y. Moreover,

$$\begin{aligned}
&\frac{\Delta}{Mh}\sum_{i=1}^{k}K^2\left(\frac{X_{i\Delta}-x}{h}\right)\underline{h}\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},y+\underline{h}z)\,dz\int_{-\infty}^{\infty}K(z)p(\underline{\Delta},X_{i\Delta},v+\underline{h}z)\,dz\\
&\qquad= O\left(\underline{h}p(\underline{\Delta},x,y)p(\underline{\Delta},x,v)\right) = o\left(p(\underline{\Delta},x,y)\right)\ \text{a.s.}
\end{aligned}$$

uniformly in $1 \le k \le n\kappa_x$, and in x, y and v. Consequently, we have from (95)

$$[Q^\Delta(x,y), Q^\Delta(x,v)]_t = o(p(\underline{\Delta}, x, y)) \tag{96}$$

uniformly in $t \le \kappa_x T$, and in x, y and v outside a neighborhood of y.

Secondly, we let $x \neq u$ and $y = v$, and consider $[Q^\Delta(x,y), Q^\Delta(u,v)]$. We have for all u outside a neighborhood of x

$$\begin{aligned}
&\frac{\Delta}{Mh}\sum_{i=1}^{k} K\left(\frac{X_{i\Delta}-x}{h}\right) K\left(\frac{X_{i\Delta}-u}{h}\right)\int_{-\infty}^{\infty} K^2(z) p(\underline{\Delta}, X_{i\Delta}, y+\underline{h}z)\,dz \\
&= \frac{\Delta}{Mh}\sum_{i=1}^{k} K\left(\frac{X_{i\Delta}-x}{h}\right) K\left(\frac{X_{i\Delta}-x}{h}+\frac{x-u}{h}\right)\int_{-\infty}^{\infty} K^2(z) p(\underline{\Delta}, X_{i\Delta}, y+\underline{h}z)\,dz \\
&= o(p(\underline{\Delta}, x, y)) \text{ a.s.}
\end{aligned}$$

uniformly in $1 \le k \le \max(n\kappa_x, n\kappa_u)$, and in x and y. Furthermore,

$$\begin{aligned}
\frac{\Delta}{Mh}\sum_{i=1}^{k} K\left(\frac{X_{i\Delta}-x}{h}\right) K\left(\frac{X_{i\Delta}-u}{h}\right)\underline{h}\left(\int_{-\infty}^{\infty} K(z) p(\underline{\Delta}, X_{i\Delta}, y+\underline{h}z)\,dz\right)^2 \\
= O\left(\underline{h}p(\underline{\Delta}, x, y)p(\underline{\Delta}, u, y)\right) = o\left(p(\underline{\Delta}, x, y)\right)
\end{aligned}$$

uniformly in $1 \le k \le \max(n\kappa_x, n\kappa_u)$, and in x, y and u. Therefore, it follows that

$$[Q^\Delta(x,y), Q^\Delta(u,y)]_t = o(p(\underline{\Delta}, x, y)) \tag{97}$$

uniformly in $t \le \max(\kappa_x T, \kappa_u T)$, and in x, y and u outside a neighborhood of x.

The results in (94), (96) and (97) we obtained above for the quadratic variation and covariations for the family of continuous martingales $Q^\Delta(x,y)$ make it clear that $Q^\Delta_{\kappa_x T}(x,y)$ is asymptotically normal with variance κ_2^2 for every $(x,y) \in \mathbb{R}^2$, and that $Q^\Delta(x,y)$ and $Q^\Delta(u,v)$ are asymptotically independent for all (x,y) and $(u,v) \in \mathbb{R}^2$ such that $(x,y) \neq (u,v)$. The reader is referred to, e.g., Revuz and Yor (1994, Chapter XIII) for the asymptotic theory of continuous martingales. Due to (92), we may therefore have that

$$\left(\frac{Mh\underline{h}}{\Delta}\right)^{1/2} \tilde{q}_\Delta(\underline{\Delta}, x, y) =_d \mathbb{N}\left(0, \kappa_2^2 p(\underline{\Delta}, x, y)\right)[1 + o(1)] \text{ a.s.} \tag{98}$$

for all $x, y \in \mathbb{R}$, and that $\tilde{q}_\Delta(\underline{\Delta}, x, y)$ and $\tilde{q}_\Delta(\underline{\Delta}, u, v)$ are asymptotically independent for all $(x,y) \neq (u,v) \in \mathbb{R}^2$ in the development of our asymptotic theories given below.

Consequently, it now follows from (98) that $f_\Delta(t,z)$ has limiting normal distribution as $\Delta \to 0$. Moreover, we may also deduce from (98) that $f_\Delta(t,u)$ and $f_\Delta(s,v)$ have asymptotic variance given by

$$\begin{aligned}
\kappa_2^2 \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(\underline{\Delta}\sum_{i=0}^{[mt]} p(i\underline{\Delta}, x_0, x) p(([mt]-i)\underline{\Delta}, y, u)\right) \\
p(\underline{\Delta}, x, y)\left(\underline{\Delta}\sum_{j=0}^{[ms]} p(j\underline{\Delta}, x_0, x) p(([ms]-j)\underline{\Delta}, y, v)\right) + o(1) \text{ a.s.,}
\end{aligned} \tag{99}$$

which converges a.s. to

$$\kappa_2^2 \int_{-\infty}^{\infty} dx \left(\int_0^t p(w, x_0, x) p(t-w, x, u)\, dw \right) \left(\int_0^s p(w, x_0, x) p(s-w, x, v)\, dw \right) \qquad (100)$$

as $\Delta \to 0$. To obtain (100), note that

$$\begin{aligned} &\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(\underline{\Delta} \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) \Big| p(([mt]-k)\underline{\Delta}, y, u) - p(([mt]-k)\underline{\Delta}, x, u) \Big| \right) \\ &\qquad p(\underline{\Delta}, x, y) \left(\underline{\Delta} \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) p(([mt]-k)\underline{\Delta}, x, v) \right) \\ &\quad \le \sup_{x \in \mathbb{R}} \left(\underline{\Delta} \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) p(([mt]-k)\underline{\Delta}, x, v) \right) \\ &\qquad \omega(\underline{\Delta}) \underline{\Delta} \int_{-\infty}^{\infty} dx \left(\sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) \Big| p(([mt]-k)\underline{\Delta}, y, u) - p(([mt]-k)\underline{\Delta}, x, u) \Big| \right) \end{aligned}$$

and that

$$\begin{aligned} \underline{\Delta} \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) p(([mt]-k)\underline{\Delta}, x, v) &= \int_0^t p(s, x_0, x) p(t-s, x, v)\, ds + o(1) \text{ a.s.} \\ &\le c_0^2 \int_0^t \frac{1}{\sqrt{s(t-s)}}\, ds + o(1) \text{ a.s.} \end{aligned}$$

uniformly in t, v, x, as $\Delta \to 0$. Moreover,

$$\begin{aligned} &\int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(\underline{\Delta} \sum_{k=0}^{[mt]} p(k\underline{\Delta}, x_0, x) \Big| p(([mt]-k)\underline{\Delta}, y, u) - p(([mt]-k)\underline{\Delta}, x, u) \Big| \right) \\ &\quad p(\underline{\Delta}, x, y) \left(\underline{\Delta} \sum_{k=0}^{[ms]} p(k\underline{\Delta}, x_0, x) \Big| p(([ms]-k)\underline{\Delta}, y, v) - p(([ms]-k)\underline{\Delta}, x, v) \Big| \right) = o(1) \text{ a.s.,} \end{aligned}$$

since

$$\begin{aligned} &\sum_{k=0}^{[ms]} p(k\underline{\Delta}, x_0, x) \Big| p(([ms]-k)\underline{\Delta}, y, v) - p(([ms]-k)\underline{\Delta}, x, v) \Big| \\ &\quad \le \frac{c_0}{\underline{\Delta}^{1/2}} \sum_{k=0}^{[ms]} \Big| p(([ms]-k)\underline{\Delta}, y, v) - p(([ms]-k)\underline{\Delta}, x, v) \Big| \\ &\quad \le \frac{c_0 c_1}{\underline{\Delta}^{1/2}} \sum_{k=0}^{[ms]-1} \frac{1}{([ms]-k)\underline{\Delta}} \end{aligned}$$

The asymptotic variance given in (99) now reduces to

$$\kappa_2^2 \int_{-\infty}^{\infty} dx \int_{-\infty}^{\infty} dy \left(\underline{\Delta} \sum_{i=0}^{[mt]} p(i\underline{\Delta}, x_0, x) p(([mt]-i)\underline{\Delta}, x, u) \right)$$
$$p(\underline{\Delta}, x, y) \left(\underline{\Delta} \sum_{j=0}^{[ms]} p(j\underline{\Delta}, x_0, x) p(([ms]-j)\underline{\Delta}, x, v) \right) + o(1) \text{ a.s.},$$

and we may obtain (100) upon noticing that

$$\int_{-\infty}^{\infty} p(\underline{\Delta}, x, y)\, dy = 1.$$

Fourth Step We are now ready to prove the stated results. For the proof of the first part, we write

$$\begin{aligned} \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \left(\hat{\lambda}_\Delta(T,x) - \lambda(T,x)\right) &= \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \int_0^T [\hat{p}_\Delta(t,x_0,x) - p(t,x_0,x)]\, dt \\ &= \int_0^T f_\Delta(t,x)\, dt + o_p(1). \end{aligned}$$

As we show in the second step of the proof, it converges in distribution to normal law with variance given by

$$\begin{aligned} &\kappa_2^2 \int_0^T dt \int_0^T ds \int_{-\infty}^{\infty} dw \left(\int_0^t p(u,x_0,w) p(t-u,w,x)\, du \right) \left(\int_0^s p(u,x_0,w) p(s-u,w,y)\, du \right) \\ &\qquad = \kappa_2^2 \int_{-\infty}^{\infty} dw \left(\int_0^T dt \int_0^t ds p(s,x_0,w) p(t-s,w,x) \right)^2, \end{aligned}$$

as was to be shown.

For the proof of the second part, we write

$$\begin{aligned} V_\Delta(T,x) &= \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \left(\hat{\Lambda}_\Delta(T,x) - \Lambda(T,x)\right) \\ &= \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \int_{-\infty}^{x} \int_0^T [\hat{p}_\Delta(t,x_0,z) - p(t,x_0,z)]\, dt\, dz \\ &= \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right)^{1/2} \int_{-\infty}^{x} \int_0^T f_\Delta(t,z)\, dt\, dz \end{aligned}$$

similarly as for the proof of the first part. Now it is clear that $V_\Delta(T,x)$ is asymptotically normal, and that the asymptotic covariance between $V(T,x)$ and $V(T,y)$ is given by the limit of

$$\int_0^T dt \int_0^T ds \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta}\right) \mathbb{E} \left(\int_{-\infty}^{x} f_\Delta(t,z)\, dz \right) \left(\int_{-\infty}^{y} f_\Delta(s,z)\, dz \right),$$

which becomes

$$\kappa_2^2 \int_{-\infty}^{\infty} dw \int_0^T dt \left(\int_0^t du\, p(u, x_0, w) \int_{-\infty}^{x} dz\, p(t-u, w, z) \right)$$
$$\int_0^T ds \left(\int_0^s du\, p(u, x_0, w) \int_{-\infty}^{y} dz\, p(s-u, w, z) \right),$$

due to our previous results. This proves the weak convergence of the finite dimensional distributions of $V_\Delta(T, \cdot)$ to those of $V(T, \cdot)$.

Now it suffices to show that the distribution of a sequence of stochastic processes $V_\Delta(T, \cdot)$ is weakly relatively compact. See, e.g., Revuz and Yor (1994, Chapter 13) for more discussions on the subject. To prove the weak relative compactness, we consider

$$\begin{aligned} V_\Delta(T, x) - V_\Delta(T, y) &= \left(\frac{Mh\underline{h}\underline{\Delta}^2}{\Delta} \right)^{1/2} \int_x^y \int_0^T [\hat{p}_\Delta(t, x_0, z) - p(t, x_0, z)]\, dt \\ &= \int_x^y \int_0^T f_\Delta(t, z)\, dt\, dz + o_p(1) \text{ a.s.} \end{aligned}$$

Similarly as above, we may deduce that

$$\begin{aligned} \mathbb{E}\Big|V_\Delta(T, x) - V_\Delta(T, y)\Big|^2 &= \int_{-\infty}^{\infty} dw \int_0^T dt \left(\int_0^t dr\, p(r, x_0, w) \int_x^y dz\, p(t-r, w, z) \right) \\ &\qquad \int_0^T ds \left(\int_0^s dr\, p(r, x_0, w) \int_x^y dz\, p(s-r, w, z) \right) + o(1) \text{ a.s.} \end{aligned}$$

However, we have

$$\begin{aligned} &\int_{-\infty}^{\infty} dw \int_0^T dt \left(\int_0^t dr\, p(r, x_0, w) \int_x^y dz\, p(t-r, w, z) \right) \\ &= \int_0^T dt \int_x^y dz \int_0^t dr \left(\int_{-\infty}^{\infty} dw\, p(r, x_0, w)\, p(t-r, w, z) \right) \\ &= \int_0^T dt \int_x^y dz \int_0^t dr\, p(t, x_0, z) = \int_0^T \left(\int_x^y tp(t, x_0, z)\, dz \right) dt \\ &\leq c|x-y| \int_0^T t^{1/2} dt \end{aligned} \tag{101}$$

for some constant $c > 0$. Moreover, uniformly in $w \in \mathbb{R}$,

$$\begin{aligned} &\int_0^T ds \left(\int_0^s dr\, p(r, x_0, w) \int_x^y dz\, p(s-r, w, z) \right) \\ &\leq c|x-y| \int_0^T \int_0^s [r(s-r)]^{-1/2}\, dr\, ds \\ &= c|x-y| \int_0^T r^{-1/2} \int_r^T (s-r)^{-1/2}\, ds\, dr \\ &= 2c|x-y| \int_0^T r^{-1/2} (T-r)^{1/2}\, dr \end{aligned} \tag{102}$$

for some constant $c > 0$. Therefore, it follows from (101) and (102) that

$$\mathbb{E}\Big| V_\Delta(T,x) - V_\Delta(T,y)\Big|^2 \le c(x-y)^2$$

for all Δ small, where $c > 0$ is some constant. This, together with the fact that $V_\Delta(T,-\infty) = 0$ a.s., establishes the weak relative compactness of the distribution of $V_\Delta(T,\cdot)$, due to the Kolmogorov's criterion for weak compactness in, e.g., Revuz and Yor (1994, Theorem 1.8, p489). The proof is therefore complete. □

Proof of Corollary 5.6

The proof is quite involved, but rather similar to that of Theorem 5.5. Whenever possible, we will therefore simply refer to the corresponding part of the proof of Theorem 5.5 without providing details. Throughout the proof, we denote by $x = (x_1, x_2)$ and $y = (y_1, y_2)$ and let

$$\begin{aligned} p_1^{X,Y}(t,x,y_1) &= \int_{-\infty}^{\infty} p^{X,Y}(t,x,(y_1,y_2))\,dy_2 \\ p_2^{X,Y}(t,x,y_2) &= \int_{-\infty}^{\infty} p^{X,Y}(t,x,(y_1,y_2))\,dy_1. \end{aligned}$$

As in the proof of Theorem 5.5, we first note

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y_1}{\underline{h}}\right)\Big| X_{i\Delta}=x_1, Y_{i\Delta}=x_2\right] \\ &= \frac{1}{\underline{h}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{z_1-y_1}{\underline{h}}\right)p^{X,Y}(\underline{\Delta},x,(z_1,z_2))\,dz_1dz_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(z_1)p^{X,Y}(\underline{\Delta},x,(y_1+\underline{h}z_1,z_2))\,dz_1dz_2 \\ &= \int_{-\infty}^{\infty} K(z_1)p_1^{X,Y}(\underline{\Delta},x,y_1+\underline{h}z_1)\,dz_1 \end{aligned}$$

and decompose

$$\begin{aligned} \frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y_1}{\underline{h}}\right) &= \int_{-\infty}^{\infty} K(z_1)p_1^{X,Y}(\underline{\Delta},(X_{i\Delta},Y_{i\Delta}),y_1+\underline{h}z_1)\,dz_1 \\ &\quad + \left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y_1}{\underline{h}}\right) - \int_{-\infty}^{\infty} K(z_1)p_1^{X,Y}(\underline{\Delta},(X_{i\Delta},Y_{i\Delta}),y_1+\underline{h}z_1)\,dz_1\right]. \end{aligned}$$

By the same token, we may also deduce that

$$\begin{aligned} &\mathbb{E}\left[\frac{1}{\underline{h}}K\left(\frac{Y_{\underline{\Delta}+i\Delta}-y_2}{\underline{h}}\right)\Big| X_{i\Delta}=x_1, Y_{i\Delta}=x_2\right] \\ &= \frac{1}{\underline{h}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{z_2-y_2}{\underline{h}}\right)p^{X,Y}(\underline{\Delta},x,(z_1,z_2))\,dz_1dz_2 \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(z_2)p^{X,Y}(\underline{\Delta},x,(z_1,y_2+\underline{h}z_2))\,dz_1dz_2 \\ &= \int_{-\infty}^{\infty} K(z_2)p_2^{X,Y}(\underline{\Delta},x,y_2+\underline{h}z_2)\,dz_2 \end{aligned}$$

with the corresponding decomposition

$$\frac{1}{\underline{h}}K\left(\frac{Y_{\underline{\Delta}+i\Delta}-y_2}{\underline{h}}\right) = \int_{-\infty}^{\infty} K(z_2)p_2^{X,Y}(\underline{\Delta},(X_{i\Delta},Y_{i\Delta}),y_2+\underline{h}z_2)\,dz_2 + \left[\frac{1}{\underline{h}}K\left(\frac{Y_{\underline{\Delta}+i\Delta}-y_2}{\underline{h}}\right) - \int_{-\infty}^{\infty} K(z_2)p_2^{X,Y}(\underline{\Delta},(X_{i\Delta},Y_{i\Delta}),y_2+\underline{h}z_2)\,dz_2\right]$$

similarly as above.

Exactly as in the second step of the proof of Theorem 5.5, we may show that the trend parts of the above decompositions are asymptotically negligible. Moreover, the martingale parts of the above decompositions can be analyzed as in the third step of the proof of Theorem 5.5. In particular, we define continuous martingales $Q_t^{\Delta,X}(x_1,y_1)$ and $Q_t^{\Delta,Y}(x_2,y_2)$ similarly as $Q_t^{\Delta}(x,y)$ introduced in the third step of the proof of Theorem 5.5. It is straightforward to obtain their quadratic covariation, though the required derivation is lengthy and tedious. As can be easily seen in the proof of Theorem 5.5, their quadratic covariation is essentially determined by

$$\begin{aligned}
&\underline{h}\mathbb{E}\left[\frac{1}{\underline{h}}K\left(\frac{X_{\underline{\Delta}+i\Delta}-y_1}{\underline{h}}\right)\frac{1}{\underline{h}}K\left(\frac{Y_{\underline{\Delta}+i\Delta}-y_2}{\underline{h}}\right)\middle| X_{i\Delta}=x_1, Y_{i\Delta}=x_2\right] \\
&= \underline{h}\left[\frac{1}{\underline{h}^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K\left(\frac{z_1-y_1}{\underline{h}}\right)K\left(\frac{z_1-y_1}{\underline{h}}\right)p^{X,Y}(\underline{\Delta},x,z)\,dz_1dz_2\right] \\
&= \underline{h}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} K(z_1)K(z_2)p^{X,Y}(\underline{\Delta},x,y+hz)\,dz_1dz_2\right],
\end{aligned}$$

whose leading term is given by

$$\underline{h}p^{X,Y}(\underline{\Delta},x,y) \le c_0\underline{h}/\underline{\Delta} = o(1) \text{ a.s.}$$

uniformly in x and y.

Moreover, if we let $\kappa = \min(\kappa_x,\kappa_y)$, then we have

$$\begin{aligned}
&\left|\frac{\Delta}{Mh}\sum_{i=1}^{n\kappa}K\left(\frac{X_{i\Delta}-x_1}{\underline{h}}\right)K\left(\frac{Y_{i\Delta}-y_1}{\underline{h}}\right)\right| \\
&\quad\le \left(\max_{s\in\mathbb{R}}|K(s)|\right)\left(\frac{\Delta}{Mh}\sum_{i=1}^{n\kappa}\left|K\left(\frac{X_{i\Delta}-x_1}{\underline{h}}\right)\right|\right) = O(1) \text{ a.s.}
\end{aligned}$$

as shown in the proof of Theorem 5.5. It is now quite clear that the two continuous martingales $Q_t^{\Delta,X}(x_1,y_1)$ and $Q_t^{\Delta,Y}(x_2,y_2)$ are asymptotically independent for all x and y. The rest of the proof is identical to that of Theorem 5.5. □

Proof of Corollary 5.7 The stated results follow straightforwardly from the proofs of Theorem 5.5 and Corollary 5.6. The details are therefore omitted. □

References

Andrews, D.W.K. and D. Pollard (1994), "An introduction to functional central limit theorems for dependent stochastic processes," *International Statistical Review*, 62, 119-132.

Bandi, F. (2002). "Short-term interest rate dynamics: A spatial approach," *Journal of Financial Economics*, 65, 73-110.

Bandi, F. and P.C.B. Phillips (2003). "Fully nonparametric estimation of scalar diffusion models," *Econometrica*, 71, 241-283.

Bosq, D. (1998). *Nonparametric Statistics for Stochastic Processes*. Springer-Verlag: New York.

Carr, P.P. and R.A. Jarrow (1990). "The stop-loss start-gain paradox and option valuation: A new decomposition into intrinsic and time value," *Review of Financial Studies*, 3, 469-492.

Durrett, R. (1991). *Probability: Theory and Examples*. Wadsworth & Brooks/Cole: California.

Florens-Zmirou, D. (1993). "On estimating the diffusion coefficient from discrete observations," *Journal of Applied Probability*, 30, 790-804.

Friedman, A. (1964). *Partial Differential Equations of Parabolic Type*. Prentice Hall: New Jersey.

Hall, P. and C.C. Heyde (1980). *Martingale Limit Theory and Its Application*. Academic Press: New York.

Karatzas, I. and S.E. Shreve (1991). *Brownian Motion and Stochastic Calculus*, 2nd ed. Springer-Verlag: New York.

Linton, O., E. Maasoumi and Y. -J. Whang (2003). "Consistent testing for stochastic dominance under general sampling schemes," mimeograph, London School of Economics.

Phillips, P.C.B. (2003). "Econometric analysis of Fisher's equation," forthcoming in J. Geanakoplos and W. Brainard eds., *Essays in Memory of Irving Fisher*.

Politis, D.N., J.P. Romano and M. Wolf (1999). *Subsampling*. Springer-Verlag: New York.

Pollard, D. (1990). *Empirical Processes: Theory and Applications*, CBMS Conference Series in Probability and Statistics, vol. 2. Institute of Mathematical Statistics, Hayward.

Revuz, D. and M. Yor (1994). *Continuous Martingale and Brownian Motion*, 2nd ed. Springer-Verlag: New York.

Serfling, R.J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley: New York.