Nonparametric Estimation for Stochastic Differential Equations

M.Sc. Thesis

by

Nishi Thakur

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2024

Nonparametric Estimation for Stochastic Differential Equations

A THESIS

Submitted in partial fulfillment of the requirements for the award of the degree of

Master of Science

by

Nishi Thakur

(Roll No.2203141005)

Under the guidance of

Dr. Debopriya Mukherjee

DEPARTMENT OF MATHEMATICS

INDIAN INSTITUTE OF TECHNOLOGY INDORE

MAY 2024

INDIAN INSTITUTE OF TECHNOLOGY INDORE

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled **"NONPARAMETRIC ESTIMATION FOR STOCHASTIC DIFFERENTIAL EQUATIONS"** in the partial fulfillment of the requirements for the award of the degree of **MASTER OF SCIENCE** and submitted in the **DEPARTMENT OF MATHEMATICS**,**INIDAN INSTITUTE OF TECHNOLOGY, INDORE**, is an authentic record of my own work carried out during the time period from July 2023 to May 2024 under the supervision of **Dr. Debopriya Mukherjee**, Assistant Professor, Department of Mathematics, IIT Indore. The matter presented in this thesis by me has not been submitted for the award of any other degree of this or any other institute.

Nishi 29/05/2024

Signature of the student with date

(**Nishi Thakur**)

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Debopriya Mukherjee 30-05-2024

Signature of Thesis Supervisor with date

(**Dr. Debopriya Mukherjee**)

Nishi Thakur has successfully given her M.Sc. Oral Examination held on 27th May, 2024.

Debopriya Mukherjee

V.K. Theni

Signature of supervisor of M.Sc Thesis

Date: 30-05-2024

Signature of Convener, DPGC

Date: 30/05/2024

Dedicated to my family and teachers.

“What, like it’s hard?”

–***Elle Woods (Legally Blonde)***

Acknowledgements

I am deeply appreciative of my supervisor Dr. Debopriya Mukherjee for her assistance and guidance throughout my M.Sc. project. She has been a major source of encouragement and inspiration for me. Her patience and determination while working pushed me to work harder. I am completing my M.Sc. with much more maturity in mathematics and her role is magnificent in this. I am also thankful to her for guiding me in my personal life as well and uplifting me. I would further like to extend my gratitude to HOD Dr. Niraj Kumar Shukla. I am grateful for the guidance and support from Dr. Swadesh Kumar Sahoo who not only improved my skills in mathematics but also guided me in my spiritual life. I am also thankful to all our teachers for their kind and valuable suggestions. I am also grateful to the Department of Mathematics, IIT Indore for providing me with great facilities such as the Bhaskaracharya lab, and ensuring my comfort while conducting the research.

I would also like to thank my family as they are my backbone. I thank my father for motivating me to become a better version of myself and always reminding me that the sky is the limit. I thank my mother for having faith in me and having my back when things got tough. Though with reluctance, I would also like to thank my brother for cheering me up and always being by my side. At last, I would like to thank my friends and all the other great people that I met on this journey for their support. I am grateful for Shivansh and Shreyal, who have been my biggest supporters throughout these years. I am also grateful to Kanchan, Madhurima, Shubham, Kevin, Uttam, Krishna, and Deepak for their love and support. I would also like to give a special mention to Dr. Brijesh Kumar Jha for taking time out of his busy schedule to help me with my thesis.

Abstract

The aim of our work is to estimate the drift and diffusion coefficients of stochastic differential equations (SDE) nonparametrically using n i.i.d replicates, $\{X_i(t) : t \in [0,1]\}_{1 \le i \le n}$, which are prone to additive noise corruption and are observed sparsely and erratically on the interval [0,1]. The word "sparse" suggests that the number of measurements per path is arbitrary, possibly as low as two, and that they stay constant with respect to n. For the estimation problem, imposing the assumption of smoothness to use smoothing techniques that further annihilate noise leads to the exclusion of a range of stochastic processes including the diffusion process. However, the estimators used in this thesis allow the functional data analysis of the processes that have nowhere differentiable sample paths, even if the observations are discrete and include noise. We talk about, $dX(t) = \mu(t)(X(t))^{\alpha}dt + \sigma(t)(X(t))^{\beta}dB(t)$ where $\alpha \in \{0,1\}$ and $\beta \in \{0,1/2,1\}$. The time-inhomogeneous SDE is one way to represent this. Using systems of PDEs, the estimators have been built by connecting the local diffusion parameters to the global parameters. This approach is entirely nonparametric and is motivated by functional data analysis. The given estimators' uniform asymp-

totic convergence demonstrates how the sample frequency influences the rate of convergence.

Contents

List of Figures

CHAPTER 1

Introduction

"Some things will drop out of the public eye and will go away, but there will always be science, engineering, and technology. And there will always, always be mathematics."

-Katherine Johnson

In this chapter, we delve into the reasons why Stochastic Differential Equations (SDEs) have garnered significant interest among mathematicians, exploring the theoretical and practical importance of SDEs and highlighting their applications. Furthermore, we discuss the various approaches to inference for SDEs, emphasizing the challenges and techniques involved in parameter estimation and model validation. Finally, we detail the specific methodology used in our study which includes a step-by-step explanation of our analytical framework.

1.1 Why Stochastic Differential Equations?

Stochastic differential equations (SDEs), despite being fairly new to the world of mathematics, have proved their importance. They are useful in the realistic modeling of various real-life problems since they permit the addition of randomness in the system. The ideal conditions required for solving ordinary or partial differential equations can barely be observed in real-life scenarios and hence one needs to add noise or randomness in the system to obtain better results.
In [3] the authors construct a nonparametric estimator for state-price density implicit for option prices and further get its asymptotic sampling theory. The estimator allows us to price new or complex securities without any arbitrage while being able to observe the attributes of the data that are important from an asset-pricing perspective, like, skewness, kurtosis for asset returns, and volatility for option prices. They further perform the Monte Carlo simulations and derive the option prices.

In [10] the authors introduce a nonparametric method for estimatimating the drift and diffusion coefficients of SDEs by the use of a densely observed discrete time series. By using Gaussian processes as priors, they can work in a function-space view allowing the inference to also happen in the same space directly. For the computational complexities that come with the use of Gaussian processes, an approximation of sparse Gaussian processes is given which allows efficient computation employing the distribution on a small subset of pseudosamples. The method is used for real data thus proving its capability to capture the demeanor of complex systems.

The stochastic volatility (SV) models, despite having an inherent insight have limited applications because of the hurdles involved in their estimation with the primary problem being the evaluation of likelihood. In recent times though, many new estimation methods have been introduced and thus the literature of SV models is being studied a lot more. In [5] the authors review some of the literature and describe the main estimators along with their advantages as well as the limitations that come with each one of them. Lately, the use of SV models, Stochastic Differential Equations(SDEs), and Diffusion Processes has increased significantly to model the dynamic evolution in various systems including the phenomena occurring in domains such as finance, biology, engineering, physics, and many more. Many authors have conducted studies and proposed a new class of SV models where the volatility is transformed according to the Box-Cox power function. In the context of stochastic differential equations, asymptotic theory of nonparametric estimation for SDEs with small noise has gained increasing attention since Kutoyants (1994) discussed consistency and asymptotic normality of a nonparametric estimator for SDEs motivated by a wiener process with small noise. Other researchers such as [10] have also conducted studies using Gaussian and small noise respectively.

Through the above examples, it is not hard to see how SDEs have gained ample importance in the field of mathematics and are being rigorously studied. Intuitively, Stochastic Differential Equations blend a deterministic equation of motion with erratic fluctuations that disrupt its dynamic evolution. In the sense of an SDE, the Diffusion Process can be understood as the equation given below,

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t), \ t \in [0, T], \tag{1.1}$$

where B(t) is the standard Brownian Motion. The above integral is supposed to be construed as an Ito integral. Hence we can observe that any diffusion process consists of two elements. The (infinitesimal) conditional variance, or *diffusion* σ, is the second component, and the (infinitesimal) conditional mean, or *drift* μ, is the first. These parameters represent the probabilistic nature of the solutions.

[9] gave a nonparametric method to perform functional data analysis for sparse longitudinal data. In the method proposed, the number of measurements at hand is small in number and repeated. This method allows us to predict individual smooth trajectories despite only a few measurements being available.
In statistical inference, the estimation of drift and diffusion coefficients has been largely studied in parametric as well as nonparametric cases. For this inference problem, an interesting fact to be considered is that diffusions are discovered discretely thus leading to disparate estimation regimens pivoted on the type of asymptotics.

1.2 Inference for SDEs

This section discusses the development of inference for SDEs. The theory developed earlier for the estimation of drift in SDEs can be observed in the time-homogeneous case. The estimation was done using one sample solution and derived from Kernel methods for parametric as well as nonparametric cases. Path observations were done either continuously (see [4] and [19] among others) or discretely (see [1] and [17] among others), using the asymptotic theory of estimators for the latter case. When considering the diffusion processes, for the sake of mathematical simplification, the authors imposed the conditions of ergodicity and stationarity along with the standard asymptotic theory. In the past few years, the kernel-based approach has been broadened to a multidimensional (see [2], [18] and [15] among others) setting for the drift coefficient keeping stationarity within the setting.

Another approach used for estimation is based on using a fixed estimation set along with the group of finite dimensional subspaces of $L^2(A, dx)$ and choosing estimators based on the minimum least square difference from each subspace (see [7] and [11] among others). However, in this approach, there is a support constraint. These time-homogeneous cases can not be further extended for the study of time-varying cases. While ideally, one would prefer having a framework that is free from any restrictions, such models, in a statistical sense are seldom recognized. Thus for time-varying SDE, we require the imposition of some semiparametric form for drift and/or diffusion coefficient.

1.3 Our Methodology

We study inferences for the continuous time stochastic process by gathering n replicates of the stochastic process $\{X_i(t)\}_{i=1}^n$. We consider them to be samples of random elements of size n that are separable in the Hilbert space of functions. Due to this global view, we can make nonparametric inferences using the mean function and kernel of X. As we have n replications, also called panels or longitudinal settings, this allows the inferences to be nonparametric. For discrete observations made for $\{X_i(t)\}_{i=1}^n$ we use smoothing techniques which often come with C^2 assumptions.

We follow the steps mentioned below to conclude our results.

- Step 1: Consider the system $Y_{ij} = X_i(T_{ij}) + U_{ij},\ i = 1 \cdots n, j = 1 \cdots r$. To estimate the local latent covariance surface along with its partial derivatives, local linear smoothing techniques are commonly employed (see [8]). The advantage is that this approach is completely nonparametric. However, the methodology used by [9] demands the smooth covariance surface for consistent estimation. This requirement presents a challenge for Stochastic Differential Equations (SDEs), as the covariance function exhibits singularity along the diagonal.

  To address this issue, [13] propose a modification that involves defining the smoothing covariance on a closed lower (or upper) triangle $\triangle := \{(s, t) :$

$0 \leq s \leq t \leq 1\}$. Then by imposing some regularity conditions, we can apply the proposed methodology to estimate the covariance function consistently from sparse, i.i.d. samples having noise corruption with uniform convergence rates.

- Step 2: Though, it is not immediately obvious how global information like mean and covariance can be deciphered into the local information such as drift and diffusion. The Fokker Planck equation (see [16]) exhibits that the probability distribution of its solution is fully determined by the drift and diffusion coefficients. However, this is not true for all the cases. For the estimation, two systems of PDE have been provided that relate the drift and diffusion functionals to mean, covariance, and their derivatives explicitly.

- Step 3: The global or pooled information can be converted into local information by plugging the estimators into systems of PDEs.

In the following chapters, this has been discussed in detail.

CHAPTER 2

Preliminaries

"It's fine to work on any problem, so long as it generates interesting mathematics along the way - even if you don't solve it at the end of the day."
-Andrew Wiles

In this chapter, we have outlined the necessary background information for the problem, providing a comprehensive foundation for understanding the context and significance of the study. Along with this, we have presented key propositions and a significant theorem that are essential for advancing the study of the proposed estimators.

2.1 Basic Definitions and Results

In this section, we discuss Brownian Motion, Stochastic processes, Stochastic Differential Equations, and some of their properties along with Ito's lemma (refer [6] and [12]). Further, we have the existence and uniqueness theorem.

Definition 2.1. *Let $(\Omega, \mathcal{F}, P)$ be a probability space. A stochastic process is a measurable function $X(t, \omega)$ defined on product space $[0, \infty) \times \Omega$. In particular,*

(i) for each t, X(t,.) is a random variable,

(ii) for each $\omega, X(., \omega)$ is a measurable function (called sample path).

Definition 2.2. *A stochastic process $B(t, \omega)$ is called a Brownian motion if it satisfies the following conditions*

(i) $P\{\omega : B(0, \omega) = 0\} = 1$.

(ii) Almost all sample paths of $B(t, \omega)$ are continuous functions i.e., $P\{\omega : B(., \omega)$ is continuous$\}$=1.

(iii) For any $0 \leq s < t$, the random variable B(t)-B(s) is normally distributed with mean 0 and variance t-s, i.e., for any $a < b$,

$$P\{a \leq B(t) - B(s) \leq b\} = \frac{1}{\sqrt{2\pi(t-s)}} \int_a^b e^{\frac{-x^2}{2(t-s)}}.$$

(iv) $B(t, \omega)$ has independent increments, i.e., for any $0 \leq t_1 < t_2 < ... < t_n$ the random variables

$$B(t_1), B(t_2) - B(t_1), ..., B(t_n) - B(t_{n-1}).$$

are independent.

Remark 2.1. *A Brownian motion has the following properties,*

a) E[$B(t)^2$]=t at any time t.

b) For any $s, t \geq 0$, E[B(s)B(t)]=min$\{$s,t$\}$.

c) For fixed $t_0 \geq 0$, the process $B(t_0 + t) - B(t_0)$ is also a Brownian motion.

d) Brownian motion is nowhere differentiable.

Definition 2.3. *Let $\mathfrak{P} := \big(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}_t\}_{t \geq 0}, P\big)$ be a filtered probability space satisfying the usual conditions i.e.,* *(i) P is complete on $(\Omega, \mathcal{F})$,*

(ii) for each $t \geq 0$, $\mathcal{F}_t$ contains all measurable null sets,

(iii) the filtration $\mathbb{F}$ is right-continuous.

Definition 2.4. *A differential equation of the form*

$$dX(t) = b(t, X(t))\, dt + \sigma(t, X(t)) dB(t), \tag{2.1}$$

where B(t) is one-dimensional standard Brownian motion and the functions b and σ are real functions is known as stochastic differential equation.

Definition 2.5. *An n-dimensional stochastic process $\{M_t\}_{t \geq 0}$ on a filtered probability space $(\omega, \mathcal{F}, P)$ is called a martingale with respect to a filteration $\{\mathcal{M}_t\}_{t \geq 0}$ (and with respect to P) if*

(i) $\{M_t\}$ is $\{\mathcal{M}_t\}$-measurable for all t,

(ii) $E[|M_t|] \leq \infty$ for all t and,

(iii) $E[M_s | \mathcal{M}_t] = M_t$ for all $s \geq t$.

Theorem 2.1 (Existence and Uniqueness [16])**.** *Let $T > 0$ and $b(.,.) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^n, \sigma(.,.) : [0, T] \times \mathbb{R}^n \to \mathbb{R}^{n \times m}$ be measurable functions satisfying*

$$|b(t, x)| + |\sigma(t, x)| \leq C(1 + |x|); x \in \mathbb{R}^n, t \in [0, T]$$

for some constant C, $(where |\sigma|^2 = \sum |\sigma_{ij}|^2)$ and such that

$$|b(t, x) - b(t, y)| + |\sigma(t, x) - \sigma(t, y)| \leq D|x - y|;\ \ x, y \in \mathbb{R}^n, t \in [0, T]$$

for some constant D. Let Z=X(0) be the random variable which is independent of the σ-algebra generated by $B_s(.), s \geq 0$ and such that

$$E[|Z|^2] < \infty.$$

Then the stochastic differential equation

$$dX(t) = b(t, X(t))\, dt + \sigma(t, X(t)) dB(t), 0 \leq t \leq T, X(0) = Z$$

has a unique t-continuous solution $X(t, \omega)$.

Theorem 2.2 (Dominated Convergence Theorem)**.** *Let (f_n) be a sequence of complex-valued measurable functions on a measure space $(\Omega, \mathcal{F}, P)$. If the sequence converges pointwise to a function f and is dominated by some Lebesgue integrable function g in a sense that,*

$$|f_n(x)| \leq g(x)$$

for all numbers n in the index and all points $x \in \Omega$. Then, f is integrable and

$$\lim_{n\to\infty} \int_\Omega |f_n - f| \, dP = 0.$$

Theorem 2.3 (Fubini Theorem)**.** *Suppose A and B are complete measurable spaces and f(x,y) is $A \times B$ measurable. If*

$$\int_{A\times B} |f(x,y)| \, d(x,y) < \infty,$$

where the integral is taken with respect to product measure on the space over $A \times B$, then

$$\int_A \Big(\int_B f(x,y) \, dy \Big) \, dx = \int_B \Big(\int_A f(x,y) \, dx \Big) \, dy = \int_{A\times B} f(x,y) \, d(x,y)$$

the first two integrals being iterated integrals with respect to two measures, respectively, and the third being an integral with respect to a product of these two measures.

In 1944, Kiyosi Itô published a six-page paper introducing Itô calculus in which he gave the parallel for the chain rule of the Newton-Leibniz calculus for random functions like Brownian motion. Since the Brownian motion moves so rapidly that it is nowhere differentiable, we cannot apply the chain rule. Itô formula also has an additional term which comes because of the Brownian motion's quadratic variation that is nonzero. The formula is stated below.

Lemma 2.4 (Itô Lemma)**.** *Consider the SDE*

$$dX(t) = b(t, X(t)) \, dt + \sigma(t, X(t)) dB(t)$$

where X(t) is an Itô process and B(t) is Brownian motion. Let $F(t,x)$ be a real-valued function with continuous partial derivatives $F'_t(t,x)$, $F'_x(t,x)$ and $F''_{xx}(t,x)$

for all $t \geq 0$ *and* $x \in \mathcal{R}$*. Then* $F(t,x)$ *is an Itô process such that*

$$dF(t,X(t)) = \Big(F_t'(t,X(t)) + F_x'(t,X(t))\mu(t) + \frac{F_{xx}''(t,X(t))\sigma^2(t)}{2}\Big)dt$$
$$+F_x'(t,X(t))\sigma(t)dB(t). \qquad (2.2)$$

2.2 Basics of Inference

In this subpart, we talk about Local Polynomial smoothing (see [8]) which is a type of estimation technique. Examine the regression function 'm', which may be locally estimated using Taylor's expansion for z in the vicinity of x as follows.

$$m(z) \approx \sum_{i=1}^{n} m(z) = m^{(j)}(x)(z-x)^j/j! \equiv \sum_{j=0}^{p} \beta_j (z-x)^j. \qquad (2.3)$$

Hence we use the locally weighted polynomial regression

$$\sum_{i=1}^{n}(Y_i - \sum_{j=0}^{p} \beta_j (X_i - x)^j)^2 K_h(X_i - x) \qquad (2.4)$$

where K(.) denotes a kernel function and h is the bandwidth.

This method has various advantages over other methods such as Nadaraya-Watson and Gasser-Miller estimators as it requires a small degree of local polynomial thus avoiding over-parametrization. It can also be easily fused with global parametric fir to reduce the bias. Let $\hat{\beta}_j$ be the minimizer of (2.4) $\forall j = 0, 1, 2, ...$ then the estimator of $m_\nu(x)$ is

$$\hat{m}_\nu(x) = \nu! \hat{\beta}_\nu. \qquad (2.5)$$

X_i in the above equation are response variables and Y_i are covariate. K(.) is the kernel function which is a unimodal nonnegative function used to assign weight $K\{(X_i - x)/h\}$ to (X_i, Y_i) and h is the bandwidth which is the smallest number among $|X_k - X_j|$ $\forall j = 1, 2, ..n$. We perform the estimation following the method explained.

From the n i.i.d. observations consider the i^{th} observation (X_i, Y_i). Now, to each

X_i assign the weight

$$K_i(X_k) = K\{h_k^{-1}(X_i - X_k)\}. \tag{2.6}$$

Use them in the initial locally weighted polynomial regression

$$\sum_{i=1}^{n}(Y_i - \sum_{j=0}^{p}\beta_j(X_i - x)^j)^2 K_h(X_i - x). \tag{2.7}$$

Denote the fitted value of Y_k as $\hat{Y_k}$ which is $\sum_{i=1}^{n} wi(X_k)Y_i$. We define the residuals r_k as $Y_k - \hat{Y_k}$, k=1,2,..,n and robustness weights as $\delta_i = B(r_i/6M)$ where $B(t) = (1-|t|^2)^2 I_{[-1,1]}(t)$ and M is the median of $|r_1|, ..., |r_n|$.

Perform the second iteration and compute the new fitted value for $\hat{(Y_k)}$ using the weight $\delta_i K_i(X_k)$.

After N iterations, we get the locally weighted regression estimator.

2.2.1 Framework for Local Polynomial Regression

For the bivariate data $(X_1, Y_1), ..., (X_n, Y_n)$, i.i.d. sample, we wish to estimate regression function $m(x_0) = \mathbf{E}(Y|X = x_0)$ and its derivatives. The model used is

$$Y = m(X) + \sigma(X)\epsilon \tag{2.8}$$

where $\mathbf{E}(\epsilon) = 0, Var(\epsilon) = 1$.

Take the Taylor's expansion of m(x) at x_0,

$$m(x) \approx m(x_0) + m'(x_0)(x - x_0) + m''(x_0)(x - x_0)/2! + ... + m^p(x_0)(x - x_0)/p!. \tag{2.9}$$

This polynomial is fitted locally by weighted least square regression problem where we wish to minimize

$$\sum_{i=1}^{n}(Y_i - \sum_{j=0}^{p}\beta_j(X_i - x)^j)^2 K_h(X_i - x) \tag{2.10}$$

and let $\hat{\beta(j)}$ be the solution of the above equation called the minimizer. From Taylor's expansion we can see that $\hat{m_\nu}(x) = \nu!\hat{\beta}_\nu$. Now we can solve the above equation for all x_0 in the domain of interest.

Converting everything to matrix notation gives us the following $\begin{bmatrix} 1 & (X_1 - x_0) & . & . & . & (X_1 - x_0)^p \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ 1 & (X_n - x_0) & . & . & . & (X_n - x_0)^p \end{bmatrix}$ denoted by X. And let y and $\hat{\beta}$ be $\begin{bmatrix} Y_1 \\ . \\ . \\ Y_n \end{bmatrix}$ and $\begin{bmatrix} \hat{\beta}_0 \\ . \\ . \\ \hat{\beta}_p \end{bmatrix}$ respectively.

Let W be $n \times n$ diagonal matrix of weights, W=diag$\{K_h(X_i - x_0)\}$ then equation (2.10) becomes

$$\min(y - X\beta)^T \times W(y - X\beta) \quad \forall \beta = (\beta_0, ..., \beta p)^T \tag{2.11}$$

and the solution is given by $\hat{\beta} = (X^T \times W \times X)^{-}1 \times X^T \times Wy$.

There are, however, some drawbacks to using this technique. Firstly, a large bandwidth value under-parametrizes the regression causing large modeling bias, whereas a small bandwidth value over-parametrizes the unknown function causing noise in the estimate. It is recommended to use a polynomial of lowest odd order; p=$\nu + 1$ or p=$\nu + 3$. And lastly, the choice of kernel function also matters as no negative weight K should be assigned to the random variables.

CHAPTER 3

Nonparametric Estimation for Stochastic Differential Equations

"It seems to me that the poet must see what others do not see, must see more deeply than other people. And the mathematician must do the same."
-Sofia Kovalevskaya

In this chapter, we have outlined the necessary background information for the problem, along with key propositions and a significant theorem essential for advancing the study of the proposed estimators.

3.1 Background

Here, we will be engaging with the solutions of the linear SDE:

$$dX(t) = \mu(t, X(t))dt + \sigma(t, X(t))dB(t),\ t \in [0, 1], \tag{3.1}$$

or

$$X(t) = X(0) + \int_0^t \mu(s)X(s)\, ds + \int_0^t \sigma(s)X(s)dB(s),\ t \in [0, 1], \tag{3.2}$$

equivalently.

The methods used can be broadened to the diffusion process configured as

$$dX(t) = \mu(t)(X(t))^{\alpha}dt + \sigma(t)(X(t))^{\beta}dB(t), t \in [0, 1], \tag{3.3}$$

to cover a wider class of diffusion processes.

The following statements are being made under the assumption that $\mu(t)$ and $\sigma(t)$ are continuously differentiable in the unit interval. Then consider the SDEs,

$$dX_i(t) = \mu(t)X_i(t)\, dt + \sigma(t)X_i(t)\, dB_i(t). \tag{3.4}$$

where $\{X_i(t), t \in [0, T]\}$, are n iid diffusions conforming to the SDEs, $i = 1, \cdots, n$ and $\{B_i\}_{1 \le i \le n}$ is a sequence of totally independent Brownian motions and $X_1(0), \cdots, X_n(0)$ are random initial points.

Further, we have n observations as

$$Y_i(T_{ij}) = X_i(T_{ij}) + U_{ij} \quad i = 1, \cdots, n, j = 1, \cdots, r(n), \tag{3.5}$$

where

- $\{T_{ij}\}$ is the triangular arrangement of randomly chosen design points.
- $\{U_{ij}\}$ is an array of i.i.d. centred measurement errors with finite variance ν^2.
- $\{r(n)\}$ is the sequence of grid sizes, which gives the denseness of the sampling scheme. The sequence of grid sizes must be at least 2.
- $\{X_i\}$, $\{T_{ij}\}$ and $\{U_{ij}\}$ are totally independent across all indices i and j.

We are interested in estimating the drift and diffusion coefficients as they dictate the local behavior of processes. One can understand them as conditional mean and conditional variance respectively. For the stochastic differential equations that are time-invariant featuring drift and diffusion that remain constant over time, this becomes readily apparent. We have,

$$\mu(x) = \lim_{h \to 0} \mathbb{E}[X_{(t+h)} - x | X_{(t)} = x]/h, \tag{3.6}$$

$$\sigma^2(x) = \lim_{h \to 0} \mathbb{E}[(X_{(t+h)} - x)^2 | X_{(t)} = x]/h. \tag{3.7}$$

The method for estimation followed further comes from [13].

3.2 Other Technical Proofs

In this section, we discuss some propositions and their proofs which provide us with the systems of PDE to globally estimate the drift and diffusion coefficients.

Proposition 3.1. *For $\alpha = 1$ and $\beta = 0$. Let $\mu, \sigma \in C^d([0,1])$, $d \geq 1$. Then $m, D \in C^{d+1}([0,1])$ satisfies the differential equations*

$$\frac{d}{dt} m(t) = \mu(t) m(t),\ m(0) = m_0, \tag{3.8}$$

$$\frac{d}{dt} D(t) = 2\mu(t) D(t) + \sigma^2(t),\ D(0) = \mathbb{E}[X^2(0)]. \tag{3.9}$$

Proof. The considered linear SDE is given by

$$dX(t) = \mu(t) X(t)\, dt + \sigma(t)\, dW(t) \tag{3.10}$$

or equivalently,

$$X(t) = X_0 + \int_0^t \mu(s) X(s)\, ds + \int_0^t \sigma(s)\, dW(s),\ t \in [0,1]. \tag{3.11}$$

We now apply Itô formula to $\Phi(x,t) = x$ for the process (3.11). Then by taking expectation on both sides, we get

$$\mathbb{E}[X(t)] = \mathbb{E}[X(0)] + \mathbb{E}\Big[\int_0^t \mu(s) X(s)\, ds\Big] + \mathbb{E}\Big[\int_0^t \sigma(s) dB(s)\Big].$$

Using the martingale property of Brownian motion, $\{B(t)\}_{t \geq 0}$ we have

$$\mathbb{E}\Big[\int_0^t \sigma(s) dB(s)\Big] = 0.$$

Using the Fubini Theorem, one can achieve

$$m(t) = m_0 + \int_0^t \mu(s) m(s)\, ds \tag{3.12}$$

which implies

$$\frac{dm}{dt}(t) = \mu(t)m(t), \quad m(0) = m_0. \tag{3.13}$$

Now, for the second equation, apply Itô formula to $\phi(x,t) = x^2$ for the process (3.11), then for $t \in [0,T]$ we have,

$$X^2(t) = X^2(0) + 2\int_0^t X(s)\Big(\mu(s)X(s)\,ds + \sigma(s)\,dB(s)\Big) + \frac{1}{2}\int_0^t 2\sigma^2(s)\,ds. \tag{3.14}$$

It can be written as,

$$X^2(t) = X^2(0) + 2\int_0^t X(s)X(s)\mu(s)\,ds + 2\int_0^t X(s)\sigma(s)\,dB(s)\Big) + \int_0^t \sigma^2(s)\,ds. \tag{3.15}$$

Again, taking expectation on both sides and using the martingale property of Brownian motion $\{B(t)\}_{t\geq 0}$ we have

$$\mathbb{E}\Big[X^2(t)\Big] = \mathbb{E}\Big[X^2(0)\Big] + 2\int_0^t \mu(s)\,\mathbb{E}\Big[X(s-)X(s)\Big]\,ds + 2\,\underbrace{\mathbb{E}\Big[\int_0^t X(s)\,\sigma(s)\,dB(s)\Big]}_{=0}$$
$$+ \int_0^t \sigma^2(s)\,ds. \tag{3.16}$$

This yields

$$D(t) = D(0) + 2\int_0^t \mu(s)D(s)\,ds + \int_0^t \sigma^2(s)\,ds.$$

which gives,

$$\frac{d}{dt}D(t) = 2\mu(t)D(t) + \sigma^2(t), \quad D(0) = \mathbb{E}[X^2(0)]. \tag{3.17}$$

□

The above proof is for a particular case of equation (3.2) where $\alpha = 1$ and $\beta = 0$. For $\alpha \in \{0,1\}$ and $\beta \in \{0, 1/2, 1\}$ the results are listed below.

- For $\alpha = 1$ and $\beta = 1/2$, we get

$$\frac{d}{dt}m(t) = \mu(t)m(t),\ m(0) = m_0,$$
$$\frac{d}{dt}D(t) = 2\mu(t)D(t) + \sigma^2(t)m(t),\ D(0) = \mathbb{E}[X^2(0)].$$

- For $\alpha = 1$ and $\beta = 1$,

$$\frac{d}{dt}m(t) = \mu(t)m(t),\ m(0) = m_0,$$

$$\frac{d}{dt}D(t) = 2\mu(t)D(t) + \sigma^2(t)D(t),\ D(0) = \mathbb{E}[X^2(0)].$$

- For $\alpha = 0$ and $\beta = 0$,

$$\frac{d}{dt}m(t) = \mu(t),\ m(0) = m_0,$$

$$\frac{d}{dt}D(t) = 2\mu(t)m(t) + \sigma^2(t),\ D(0) = \mathbb{E}[X^2(0)].$$

- For $\alpha = 0$ and $\beta = 1/2$,

$$\frac{d}{dt}m(t) = \mu(t),\ m(0) = m_0,$$

$$\frac{d}{dt}D(t) = 2\mu(t)m(t) + +\sigma^2(t)m(t),\ D(0) = \mathbb{E}[X^2(0)].$$

- For $\alpha = 0$ and $\beta = 1$,

$$\frac{d}{dt}m(t) = \mu(t),\ m(0) = m_0,$$

$$\frac{d}{dt}D(t) = 2\mu(t)m(t) + +\sigma^2(t)D(t),\ D(0) = \mathbb{E}[X^2(0)].$$

Proposition 3.2. *Let $\mu, \sigma \in C^d([0,1])$ for some $d \geq 1$. Let $H : \Delta \to \mathbb{R}$ be given by $H(s,t) = \mathbb{E}[X(s)X(t)]$ for $(s,t) \in \Delta = \{(s,t) : 0 \leq s \leq t \leq 1\}$. Then, $H \in C^{d+1}(\Delta; \mathbb{R})$. Furthermore, for $0 \leq s \leq t \leq 1$,*

$$H(s,t) = H(0,0) + 2\int_0^s H(r,r)\mu(r)\,dr + \int_s^t H(s,r)\mu(r)\,dr + \int_0^s \sigma^2(r)\,dr. \tag{3.18}$$

In particular, for $0 \leq s \leq t \leq 1$,

$$\partial_s H(s,t) = \mu(s)H(s,s) + \int_s^t \mu(r)\partial_s H(s,r)\,dr + \sigma^2(s) \tag{3.19}$$

and the following system of PDE holds,

$$\partial m(s) = m(s)\mu(s) \tag{3.20}$$

$$\sigma^2(s) = \frac{1}{1-s}\int_s^1 \Big[\partial_s H(s,t) - \mu(s)H(s,s) - \int_s^t \mu(r)\partial_s H(s,r)\,dr\Big]\,dt. \tag{3.21}$$

Proof. The process $\{X(t)\}_{t\in[0,1]}$ satisfies

$$X(t) = X(s) + \int_s^t \mu(r)X(r)\,dr + \int_s^t \sigma(r)\,dW(r). \tag{3.22}$$

The random variables $X(t) - X(s)$ and $X(s)$ are independent for $0 \leq s \leq t$. Hence, with $D(s) = \mathbb{E}[X^2(s)]$ and $m(s) = \mathbb{E}[X(s)]$ we achieve

$$\begin{aligned} H(s,t) = \mathbb{E}[X(s)X(t)] &= \mathbb{E}\big[X(s)\big(X(t) - X(s)\big)\big] + \mathbb{E}[X^2(s)] \\ &= m(s)\mathbb{E}\big[X(t) - X(s)\big] + D(s). \end{aligned} \tag{3.23}$$

Taking expectation on (3.22) we achieve

$$\mathbb{E}\big[X(t) - X(s)\big] = \int_s^t \mu(r)\mathbb{E}[X(R)]\,dr. \tag{3.24}$$

The random variables $X(s) - X(0)$ and $X(r)$ are independent for $0 \leq s \leq r \leq t$. Also, the random variables $X(0)$ and $X(r)$ are independent for $0 \leq r$. Thus, we get

$$\begin{aligned} \mathbb{E}[X(s)]\mathbb{E}[X(r)] &= \mathbb{E}\Big[X(s) - X(0)\Big]\mathbb{E}[X(r)] + \mathbb{E}[X(0)]\mathbb{E}[X(r)] \\ &= \mathbb{E}\Big[(X(s) - X(0))X(r)\Big] + \mathbb{E}[X(0)\,X(r)] \\ &= \mathbb{E}\Big[X(s)X(r) - X(0)X(r) + X(0)X(r)\Big] \\ &= \mathbb{E}\Big[X(s)X(r)\Big] = H(s,r). \end{aligned} \tag{3.25}$$

Using Proposition 3.1 and the above equation, (3.25) we have

$$\begin{aligned} H(s,t) &= \mathbb{E}[X(s)]\Big(\int_s^t \mu(r)\mathbb{E}[X(r)]\,dr\Big) + D(s) \\ &= \int_s^t \mu(r)\mathbb{E}[X(s)]\mathbb{E}[X(r)]\,dr + D(0) + 2\int_0^s \mu(r)\,D(r)\,dr + \int_0^s \sigma^2(r)\,dr \\ &= H(0,0) + 2\int_0^s \mu(r)H(r,r)\,dr + \int_s^t \mu(r)H(s,r)\,dr \\ &= H(0,0) + 2\int_0^s \mu(r)H(r,r)\,dr + \int_s^t \mu(r)H(s,r)\,dr + \int_0^s \sigma^2(r)\,dr. \end{aligned} \tag{3.26}$$

Differentiating (3.26) with respect to s and using the Dominated Convergence theorem, we get

$$\partial_s H(s,t) = 2\mu(s)H(s,s) + \int_s^t \mu(r)\partial_s H(s,r)\,dr - \mu(s)H(s,s) + \sigma^2(s). \tag{3.27}$$

Now integrating (3.27) with respect to t variable in $[s,1]$, we obtain

$$\int_s^1 \Big[\partial_s H(s,t) - \mu(s)H(s,s) - \int_s^t \mu(r)\partial_s H(s,r)\,dr\Big]\,dt = (1-s)\Big(\sigma^2(s)\Big). \quad (3.28)$$

That is, for $0 \le s \le 1$, we obtain

$$\sigma^2(s) = \frac{1}{1-s}\int_s^1 \Big[\partial_s H(s,t) - \mu(s)H(s,s) - \int_s^t \mu(r)\partial_s H(s,r)\,dr\Big]\,dt. \quad (3.29)$$

This completes the proof. □

The above proof is for a particular case of equation (5) where $\alpha = 1$ and $\beta = 0$. For $\alpha \in \{0,1\}$ and $\beta \in \{0, 1/2, 1\}$ the results are listed below.

- For $\alpha = 1$ and $\beta = 1/2$, we get

$$\partial_s H(s,t) = \mu(s)H(s,s) + \int_s^t \mu(r)\partial_s H(s,r)\,dr + \sigma^2(s)m(s),$$

  and we get the system of PDE,

$$\partial m(s) = m(s)\mu(s)$$

$$\sigma^2(s) = \frac{1}{(1-s)m(s)}\int_s^1 \Big[\partial_s H(s,t) - \mu(s)H(s,s) - \int_s^t \mu(r)\partial_s H(s,r)\,dr\Big]\,dt.$$

- For $\alpha = 1$ and $\beta = 1$,

$$\partial_s H(s,t) = \mu(s)H(s,s) + \int_s^t \mu(r)\partial_s H(s,r)\,dr + \sigma^2(s)H(s,s),$$

  and we get the system of PDE,

$$\partial m(s) = m(s)\mu(s)$$

$$\sigma^2(s) = \frac{1}{(1-s)H(s,s)}\int_s^1 \Big[\partial_s H(s,t) - \mu(s)H(s,s) - \int_s^t \mu(r)\partial_s H(s,r)\,dr\Big]\,dt.$$

- For $\alpha = 0$ and $\beta = 0$,

$$\partial_s H(s,t) = \mu(s)m(s) + \int_s^t \mu(r)\partial_s m(s)\,dr + \sigma^2(s),$$

  and we get the system of PDE,

$$\partial m(s) = \mu(s)$$

$$\sigma^2(s) = \frac{1}{1-s}\int_s^1 \left[\partial_s H(s,t) - \mu(s)H(s,s) - \int_s^t \mu(r)\partial_s H(s,r)\,dr\right] dt.$$

- For $\alpha = 0$ and $\beta = 1/2$,

$$\partial_s H(s,t) = \mu(s)m(s) + \int_s^t \mu(r)\partial_s m(s)\,dr + \sigma^2(s)m(s),$$

  and we get the system of PDE,

$$\partial m(s) = \mu(s)$$

$$\sigma^2(s) = \frac{1}{(1-s)m(s)}\int_s^1 \left[\partial_s H(s,t) - \mu(s)m(s) - \int_s^t \mu(r)\partial_s m(s)\,dr\right] dt.$$

- For $\alpha = 0$ and $\beta = 1$,

$$\partial_s H(s,t) = \mu(s)m(s) + \int_s^t \mu(r)\partial_s m(s)\,dr + \sigma^2(s)H(s,s),$$

  and we get the system of PDE,

$$\partial m(s) = \mu(s)$$

$$\sigma^2(s) = \frac{1}{(1-s)H(s,s)}\int_s^1 \left[\partial_s H(s,t) - \mu(s)m(s) - \int_s^t \mu(r)\partial_s m(s)\,dr\right] dt.$$

3.3 Estimators

We want to estimate the mean and the second-moment function, m and G respectively, and their derivatives using local polynomial smoothing. Using all the observations, we can perform estimation by pooling second moments on the triangle,

$$\triangle := \{(s,t) : 0 \le s \le t \le 1\} \tag{3.30}$$

Using this information, one can see that for any t such that $0 \le t \le 1$ the pointwise estimator of order d obtained using local polynomial for $m(t)$ and its derivative $\partial m(t)$ are:

$$\left(\hat{m}(t), h_m\widehat{\partial m}(t)\right)^T = \left((1, \underbrace{0,...,0}_{=\text{d times}})^T\left(\hat{\beta}_p\right)_{0\le p\le d}, (0,1, \underbrace{0,..,0}_{=\text{d-1 times}})^T\left(\hat{\beta}_p\right)_{0\le p\le d}\right)^T \tag{3.31}$$

where the vector $\left(\hat{\beta}_p\right)_{0\leq p\leq d}$ is the solution of

$$\underset{(\beta_p)0\leq p\leq d}{\arg\min} \underbrace{\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(Y_{ij} - \sum_{p=0}^{d}\beta_p(T_{ij}-t)^p)^2 K_{h_m^2}(T_{ij}-t)}_{=A}, \tag{3.32}$$

h_m is a one-dimensional bandwidth parameter that depends on n and $K_{h_m^2} = {h_m}^{-1}K_m({h_m}^{-1})$ for some kernel K_m that is univariate and integrable. Similarly, for estimating G(s,t) constrained to the lower triangle,

$\Delta = \{(s,t) : 0 \leq s \leq t \leq 1\}$, we use local surface regression approach on a 2D scatter plot given below:

$$\{((T_{ik}, T_{ij}), Y_{ij}Y_{ik}) : i = 1\cdots n, k < j\}. \tag{3.33}$$

The diagonal points have been excluded as the measurement error causes the diagonal observations to be biased.

For function G(s,t) and its partial derivative $\partial_s G(s,t)$ for $s \leq t$, we have the local smoothing of order d as follows,

$$\left(\hat{G}(s,t), h_G\widehat{\partial_s G}(s,t)\right)^T = \left((1,0,...,0)^T\left(\hat{\gamma}_{p,q}\right)_{0\leq p+q\leq d}, (0,1,0,..,0)^T\left(\hat{\gamma}_{p,q}\right)_{0\leq p+q\leq d}\right)^T \tag{3.34}$$

where the vector $(\hat{\gamma}_{p,q})^T_{0\leq p+q\leq d}$ is the minimizer of

$$\underset{\gamma_{p,q}}{\arg\min}\sum_{i\leq n}\sum_{k\leq j}\left\{Y_{ik}Y_{ij} - \sum_{0\leq(p+q)\leq d}\gamma_{p,q}h^{p+q}{}_G\left(\frac{T_{ij}-s}{h_G}\right)^p\left(\frac{T_{ik}-t}{h_G}\right)^q\right\}^2 \times K_{h_G}((T_{ij}-s),(T_{ik}-t)) \tag{3.35}$$

where ${H_G}^{1/2}$ is 2×2 bandwidth matrix which is symmetric positive definite and $K_{h_G} = |H_G|^{1/2}K_{h_G}$ for some bivariate kernel K_G.

Hence, by combining these estimators of m, G, and their derivatives, obtain two simultaneous pairs of estimators $(\hat{\mu}, \hat{\sigma}_D^2)$ and $(\hat{\mu}, \hat{\sigma}_T^2)$ for the drift and diffusion functions (subscript D and T denote the diagonal and triangular domain respectively) :

$$\begin{cases}\hat{\mu}(t) & = (\hat{m}(t))^{-1}\,\widehat{\partial m}(t)\,\mathbb{I}(\hat{m}(t)\neq 0),\\ \hat{\sigma}_D^2(t) & = \widehat{\partial D}(t) - 2\hat{\mu}(t)\hat{D}(t), t\in[0,1]\end{cases} \tag{3.36}$$

$$\begin{cases} \hat{\mu}(s) = (\hat{m}(s))^{-1}\, \widehat{\partial m}(s)\, \mathbb{I}(\hat{m}(s) \neq 0), \\ \hat{\sigma}_T^2(s) = \dfrac{1}{1-s} \int_s^1 \left[\widehat{\partial_s G}(s,t) - \hat{\mu}(s)\hat{G}(s,s) - \int_s^t \hat{\mu}(u)\widehat{\partial_s G}(s,u)\, du \right] dt, \quad s \in [0,1] \end{cases} \tag{3.37}$$

for some $t \in A$ where A is some subset of [0,1].

3.4 Main Theorems

Now, we discuss the main theorems and their proofs which complete the estimation theory (refer [14]). The theorems given below hold under the assumptions:

- **C(0)** The drift and diffusion coefficients are d-times continuously differentiable on the unit interval, i.e. $\mu(.), \sigma(.) \in C^d([0,1], \mathbb{R})$ for some $d \geq 1$.
- **C(1)** $\exists M \geq 0$ which gives $0 < \mathbb{P}(T_{ij} \in [a,b]) \leq M(b-a)$ for all i,j and $0 \leq a < b \leq 1$.
- **C(2)** $\mathbb{E}|U_{ij}{}^{\rho}| < \infty$ and $\mathbb{E}|X(0)^{\rho}| < \infty$.

Theorem 3.3. *Assume the conditions C(0), C(1) and C(2) hold for $\rho > 2$ and let $\hat{m}(.)$ and $\widehat{\partial m}(.)$ be the estimator defined in* (3.31)*. Then with probability 1.*

$$\sup |\hat{m}(t) - m(t)| = O(\mathcal{R}(n)) \tag{3.38}$$

$$\sup |\widehat{\partial m}(t) - \partial m(t)| = h_m{}^{-1} O(\mathcal{R}(n)) \tag{3.39}$$

where $\mathcal{R}(n) = \left[h_m{}^{-2} \frac{logn}{n} \left(h_m{}^2 + \frac{h_m}{r} \right) \right]^{1/2} + h_m{}^{d+1}$. Additionally if C(2) holds for $\rho > 4$ and let $\hat{G}(.)$ and $\widehat{\partial_s G}(.)$ be the estimator defined in (3.34) *satisfies with probability 1.*

$$\sup |\hat{G}(s,t) - G(s,t)| = O(\mathcal{Q}(n)) \tag{3.40}$$

$$\sup |\widehat{\partial_s G}(s,t) - \partial_s G(s,t)| = h_G{}^{-1} O(\mathcal{Q}(n)) \tag{3.41}$$

where $\mathcal{Q}(n) = \left[h_G{}^{-4} \frac{logn}{n} \left(h_G{}^4 + \frac{h_G{}^3}{r} \frac{h_G{}^2}{r^2} \right) \right]^{1/2} + h_G{}^{d+1}$.

Proof. Consider the equation (3.32) For simplicity, let us take d=2. Further, we need to minimize only A. Therefore, we have

$$\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij}-\beta_0(T_{ij}-t)^0-\beta_1(T_{ij}-t)^1-\beta_2(T_{ij}-t)^2\}^2K_{h_m^2}(T_{ij}-t)]. \quad (3.42)$$

Here three parameters are to be estimated, $\beta_0, \beta_1, \beta_2$.

Note that S=$Y_i-\hat{Y}_i$= Actual value-Predicted value. Thus, in our case, the normal equations are obtained by partially differentiating S with respect to β_0, β_1 and β_2. The equation,$\frac{\partial S}{\partial \beta_0}=0$, will give an expression for $\hat{\beta}_0$. Similarly, for $\hat{\beta}_1$ and $\hat{\beta}_2$ we have the expression $\frac{\partial S}{\partial \beta_1}=0$ and $\frac{\partial S}{\partial \beta_2}=0$ respectively, where $\hat{\beta}_0$, $\hat{\beta}_1$ and $\hat{\beta}_2$ are the solutions of these PDE.

Differentiating (3.42) with respect to $\hat{\beta}_0$ yields $\frac{\partial(3.42)}{\partial \beta_0}=0$, we get

$$2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij}-\hat{\beta}_0-\hat{\beta}_1(T_{ij}-t)-\hat{\beta}_2(T_{ij}-t)^2\}(-1)K_{h_m^2}(T_{ij}-t)]=0,$$

which is nothing but,

$$-2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij}-\hat{\beta}_0-\hat{\beta}_1(T_{ij}-t)-\hat{\beta}_2(T_{ij}-t)^2\}K_{h_m^2}(T_{ij}-t)]=0.$$

Further simplification gives

$$\begin{aligned}\hat{\beta}_0\sum_{i=1}^{n}\sum_{j=1}^{r(n)}K_{h_m^2}(T_{ij}-t) &= \sum_{i=1}^{n}\sum_{j=1}^{r(n)}Y_{ij}K_{h_m^2}(T_{ij}-t)-\hat{\beta}_1\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij}-t)\\ &\quad\times K_{h_m^2}(T_{ij}-t)-\hat{\beta}_2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij}-t)^2K_{h_m^2}(T_{ij}-t).\end{aligned}$$

Hence we get the expression,

$$\begin{aligned}\hat{\beta}_0=\frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{r(n)}K_{h_m^2}(T_{ij}-t)}\Big[&\sum_{i=1}^{n}\sum_{j=1}^{r(n)}Y_{ij}K_{h_m^2}(T_{ij}-t)\\ &-\hat{\beta}_1\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij}-t)K_{h_m^2}(T_{ij}-t)\\ &-\hat{\beta}_2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij}-t)^2K_{h_m^2}(T_{ij}-t)\Big].\end{aligned}$$

Similarly, partially differentiating (3.42) with respect to β_1 ($\frac{\partial(3.42)}{\partial\beta_1} = 0$) gives us

$$2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij} - \hat{\beta}_0 - \hat{\beta}_1(T_{ij} - t) - \hat{\beta}_2(T_{ij} - t)^2\}(-1)(T_{ij} - t)K_{h_m^2}(T_{ij} - t)] = 0,$$

which gives

$$-2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij} - \hat{\beta}_0 - \hat{\beta}_1(T_{ij} - t) - \hat{\beta}_2(T_{ij} - t)^2\}(T_{ij} - t)K_{h_m^2}(T_{ij} - t)] = 0.$$

Then, clearly

$$\begin{aligned}\hat{\beta}_1\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij} - t)^2 K_{h_m^2}(T_{ij} - t) &= \sum_{i=1}^{n}\sum_{j=1}^{r(n)} Y_{ij}(T_{ij} - t)K_{h_m^2}(T_{ij} - t) \\ &\quad -\hat{\beta}_0\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij} - t)K_{h_m^2}(T_{ij} - t) \\ &\quad -\hat{\beta}_2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij} - t)^3 K_{h_m^2}(T_{ij} - t).\end{aligned}$$

Thus, we obtain

$$\begin{aligned}\hat{\beta}_1 = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij} - t)^2 K_{h_m^2}(T_{ij} - t)}\big[&\sum_{i=1}^{n}\sum_{j=1}^{r(n)} Y_{ij}(T_{ij} - t)K_{h_m^2}(T_{ij} - t) \\ &-\hat{\beta}_0\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij} - t)K_{h_m^2}(T_{ij} - t) \\ &-\hat{\beta}_2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}(T_{ij} - t)^3 K_{h_m^2}(T_{ij} - t)\big].\end{aligned}$$

Lastly, we differentiate (3.42) with respect to β_2 ($\frac{\partial(3.42)}{\partial\beta_2} = 0$), to obtain

$$2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij} - \hat{\beta}_0 - \hat{\beta}_1(T_{ij} - t) - \hat{\beta}_2(T_{ij} - t)^2\}(-1)(T_{ij} - t)^2 K_{h_m^2}(T_{ij} - t)] = 0,$$

which is

$$-2\sum_{i=1}^{n}\sum_{j=1}^{r(n)}[\{Y_{ij} - \hat{\beta}_0 - \hat{\beta}_1(T_{ij} - t) - \hat{\beta}_2(T_{ij} - t)^2\}(T_{ij} - t)^2 K_{h_m^2}(T_{ij} - t)] = 0.$$

By simplifying, we have

$$\hat{\beta}_2 \sum_{i=1}^{n}\sum_{j=1}^{r(n)} (T_{ij}-t)^4 K_{h_m^2}(T_{ij}-t) = \sum_{i=1}^{n}\sum_{j=1}^{r(n)} Y_{ij}(T_{ij}-t)^2 K_{h_m^2}(T_{ij}-t)$$
$$-\hat{\beta}_0 \sum_{i=1}^{n}\sum_{j=1}^{r(n)} (T_{ij}-t)^2 K_{h_m^2}(T_{ij}-t)$$
$$-\hat{\beta}_1 \sum_{i=1}^{n}\sum_{j=1}^{r(n)} (T_{ij}-t)^3 K_{h_m^2}(T_{ij}-t).$$

Thus, we get the expression,

$$\implies \hat{\beta}_2 = \frac{1}{\sum_{i=1}^{n}\sum_{j=1}^{r(n)} (T_{ij}-t)^4 K_{h_m^2}(T_{ij}-t)} \Big[\sum_{i=1}^{n}\sum_{j=1}^{r(n)} Y_{ij}(T_{ij}-t)^2 K_{h_m^2}(T_{ij}-t)$$
$$-\hat{\beta}_0 \sum_{i=1}^{n}\sum_{j=1}^{r(n)} (T_{ij}-t)^2 K_{h_m^2}(T_{ij}-t)$$
$$-\hat{\beta}_1 \sum_{i=1}^{n}\sum_{j=1}^{r(n)} (T_{ij}-t)^3 K_{h_m^2}(T_{ij}-t)\Big].$$

This gives us the solution for (3.32) for d=2 and hence the result follows. Now, consider the equation (3.35) For the sake of simplicity, let us take d=1. Then we will have the following cases.

(i) p=0, q=0

(ii) p=1, q=0

(iii) p=0, q=1

We have the expression,

$$\sum_{i\le n}\sum_{k\le j}[\{Y_{ik}Y_{ij} - \gamma_{0,0}{h_G}^0\Big(\frac{T_{ij}-s}{h_G}\Big)^0\Big(\frac{T_{ik}-t}{h_G}\Big)^0 - \gamma_{1,0}{h_G}^1\Big(\frac{T_{ij}-s}{h_G}\Big)^1\Big(\frac{T_{ik}-t}{h_G}\Big)^0$$
$$-\gamma_{0,1}{h_G}^1\Big(\frac{T_{ij}-s}{h_G}\Big)^0\Big(\frac{T_{ik}-t}{h_G}\Big)^1\}^2 K_{h_G}((T_{ij}-s),(T_{ik}-t))]. \tag{3.43}$$

Here we have three parameters to be estimated, $\hat{Y_{0,0}}$, $\hat{Y_{1,0}}$, and $\hat{Y_{0,1}}$. Continuing in a similar manner as for the previous proof, we partially differentiate (3.43) with respect to $\gamma_{0,0}$ ($\frac{\partial(3.43)}{\partial\gamma_{0,0}} = 0$), which gives

$$2\sum_{i\leq n}\sum_{k\leq j}[\{Y_{ik}Y_{ij} - \hat{\gamma}_{0,0} - \hat{\gamma}_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big) - \hat{\gamma}_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)\}$$
$$\times(-1)K_{h_G}((T_{ij}-s),(T_{ik}-t))] = 0,$$

which can be written as

$$-2\sum_{i\leq n}\sum_{k\leq j}[\{Y_{ik}Y_{ij} - \hat{\gamma}_{0,0} - \hat{\gamma}_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big) - \hat{\gamma}_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)\}$$
$$\times K_{h_G}((T_{ij}-s),(T_{ik}-t))] = 0.$$

Hence we have the equation

$$\begin{aligned}\hat{\gamma}_{0,0}\sum_{i\leq n}\sum_{k\leq j}K_{h_G}((T_{ij}-s),(T_{ik}-t)) &= \sum_{i\leq n}\sum_{k\leq j}Y_{ik}Y_{ij}K_{h_G}((T_{ij}-s),(T_{ik}-t)) \\ &-\sum_{i\leq n}\sum_{k\leq j}\hat{\gamma}_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t)) \\ &-\sum_{i\leq n}\sum_{k\leq j}\hat{\gamma}_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t)),\end{aligned}$$

which can be simplified to obtain

$$\begin{aligned}\hat{\gamma}_{0,0} = \frac{1}{\sum_{i\leq n}\sum_{k\leq j}K_{h_G}((T_{ij}-s),(T_{ik}-t))}\Big[&\sum_{i\leq n}\sum_{k\leq j}Y_{ik}Y_{ij}K_{h_G}((T_{ij}-s),(T_{ik}-t)) \\ &-\sum_{i\leq n}\sum_{k\leq j}\hat{\gamma}_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t)) \\ &-\sum_{i\leq n}\sum_{k\leq j}\hat{\gamma}_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t))\Big].\end{aligned} \tag{3.44}$$

Proceeding similarly by differentiating (3.43) with respect to $\gamma_{1,0}$ ($\frac{\partial(3.43)}{\partial\gamma_{1,0}} = 0$) yields

$$2\sum_{i\leq n}\sum_{k\leq j}[\{Y_{ik}Y_{ij} - \hat{\gamma}_{0,0} - \hat{\gamma}_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big) - \hat{\gamma}_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)\}(-1)h_G\Big(\frac{T_{ij}-s}{h_G}\Big)$$
$$\times K_{h_G}((T_{ij}-s),(T_{ik}-t))] = 0,$$

which is

$$-2\sum_{i\le n}\sum_{k\le j}[\{Y_{ik}Y_{ij}-\hat\gamma_{0,0}-\hat\gamma_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)-\hat\gamma_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)\}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)$$

$$\times K_{h_G}((T_{ij}-s),(T_{ik}-t))]=0.$$

We can write this as

$$\hat\gamma_{1,0}{h_G}^2\sum_{i\le n}\sum_{k\le j}\Big(\frac{T_{ij}-s}{h_G}\Big)^2K_{h_G}((T_{ij}-s),(T_{ik}-t))=$$

$$\sum_{i\le n}\sum_{k\le j}Y_{ik}Y_{ij}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t))-$$

$$\sum_{i\le n}\sum_{k\le j}\hat\gamma_{0,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t))-$$

$$\sum_{i\le n}\sum_{k\le j}\hat\gamma_{0,1}{h_G}^2\Big(\frac{T_{ik}-t}{h_G}\Big)\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t)),$$

and consequently,

$$\hat\gamma_{1,0}=\frac{1}{{h_G}^2\sum_{i\le n}\sum_{k\le j}\Big(\frac{T_{ij}-s}{h_G}\Big)^2K_{h_G}((T_{ij}-s),(T_{ik}-t))} \tag{3.45}$$

$$\Big[\sum_{i\le n}\sum_{k\le j}Y_{ik}Y_{ij}\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t))$$

$$-\sum_{i\le n}\sum_{k\le j}\hat\gamma_{0,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t)) \tag{3.46}$$

$$-\sum_{i\le n}\sum_{k\le j}\hat\gamma_{0,1}{h_G}^2\Big(\frac{T_{ij}-s}{h_G}\Big)\Big(\frac{T_{ik}-t}{h_G}\Big)K_{h_G}((T_{ij}-s),(T_{ik}-t))\Big].$$

Further the partial differentiation of (3.43) with respect to $\gamma_{0,1}$ ($\frac{\partial(3.43)}{\partial\gamma_{0,1}}=0$), we have

$$2\sum_{i\le n}\sum_{k\le j}[\{Y_{ik}Y_{ij}-\hat\gamma_{0,0}-\hat\gamma_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)-\hat\gamma_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)\}(-1)h_G\Big(\frac{T_{ik}-t}{h_G}\Big)$$

$$\times K_{h_G}((T_{ij}-s),(T_{ik}-t))]=0,$$

which gives

$$-2\sum_{i\le n}\sum_{k\le j}[\{Y_{ik}Y_{ij}-\hat\gamma_{0,0}-\hat\gamma_{1,0}h_G\Big(\frac{T_{ij}-s}{h_G}\Big)-\hat\gamma_{0,1}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)\}h_G\Big(\frac{T_{ik}-t}{h_G}\Big)$$

$$\times K_{h_G}((T_{ij}-s),(T_{ik}-t))]=0.$$

Simplify the above equation to get

$$\hat{\gamma}_{0,1} {h_G}^2 \sum_{i \le n} \sum_{k \le j} \Big(\frac{T_{ik}-t}{h_G}\Big)^2 K_{h_G}((T_{ij}-s),(T_{ik}-t)) =$$

$$\sum_{i \le n} \sum_{k \le j} Y_{ik} Y_{ij} h_G \Big(\frac{T_{ik}-t}{h_G}\Big) K_{h_G}((T_{ij}-s),(T_{ik}-t))$$

$$-\sum_{i \le n} \sum_{k \le j} \hat{\gamma}_{0,0} h_G \Big(\frac{T_{ik}-t}{h_G}\Big) K_{h_G}((T_{ij}-s),(T_{ik}-t)) -$$

$$\sum_{i \le n} \sum_{k \le j} \hat{\gamma}_{1,0} {h_G}^2 \Big(\frac{T_{ik}-t}{h_G}\Big)\Big(\frac{T_{ij}-s}{h_G}\Big) K_{h_G}((T_{ij}-s),(T_{ik}-t)),$$

Thus we obtain the value

$$\hat{\gamma}_{0,1} = \frac{1}{{h_G}^2 \sum_{i \le n} \sum_{k \le j} \Big(\frac{T_{ik}-t}{h_G}\Big)^2 K_{h_G}((T_{ij}-s),(T_{ik}-t))} \tag{3.47}$$

$$\Big[\sum_{i \le n} \sum_{k \le j} Y_{ik} Y_{ij} \Big(\frac{T_{ik}-t}{h_G}\Big) K_{h_G}((T_{ij}-s),(T_{ik}-t))$$

$$-\sum_{i \le n} \sum_{k \le j} \hat{\gamma}_{0,0} h_G \Big(\frac{T_{ik}-t}{h_G}\Big) K_{h_G}((T_{ij}-s),(T_{ik}-t)) \tag{3.48}$$

$$-\sum_{i \le n} \sum_{k \le j} \hat{\gamma}_{1,0} {h_G}^2 \Big(\frac{T_{ik}-t}{h_G}\Big)\Big(\frac{T_{ij}-s}{h_G}\Big) K_{h_G}((T_{ij}-s),(T_{ik}-t))\Big].$$

This gives us the solution of (3.35) for $d = 1$

Now for the equation (3.34). We have,

$$(\hat{\gamma}_{0,0}, \hat{\gamma}_{1,0}, \hat{\gamma}_{0,1})^T = {(\hat{\gamma}_{p,q})^T}_{0 \le p+q \le 1}$$

which is the same as

$$(\hat{\gamma}_{0,0}, \hat{\gamma}_{1,0}, \hat{\gamma}_{0,1}) = (\hat{\gamma}_{p,q})_{0 \le p+q \le 1}.$$

Thus we can see that the least square problem (3.35) has the solution,

$${(\hat{\gamma}_{p,q})^T}_{0 \le p+q \le 1} = {\Big({h_G}^{p+q} (\widehat{\partial_s^p t^p)^{p+q}} G(s,t)\Big)^T}_{0 \le p+q \le 1}. \tag{3.49}$$

Substituting the value p=0 and q=0 in the above equation, we get

$$(\hat{\gamma}_{0,0})^T = \Big({h_G}^0 \hat{G}(s,t)\Big)^T$$

that is

$$(\hat{\gamma}_{0,0}) = (\hat{G}(s,t))$$

and hence we obtain,

$$(\hat{\gamma}_{0,0})\hat{G}(s,t) = (\hat{\gamma}_{0,0}). \tag{3.50}$$

Further proceeding using the general formula,

$$\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}\mathbf{y} \tag{3.51}$$

which is nothing but

$$(\mathbf{X}^T\mathbf{X})^{-2}\mathbf{X}\mathbf{y} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}. \tag{3.52}$$

Then, the equation (3.49) can be written as

$$(\mathbb{T}_{(s,t)}{}^T\mathbf{W}_{(s,t)}\mathbb{T}_{(s,t)})^{-1}\mathbb{T}_{(s,t)}{}^T\mathbf{W}_{(s,t)}\mathbf{y} = \begin{bmatrix} \hat{G}(s,t) \\ h_G\widehat{\partial_s G}(s,t) \\ h_G\widehat{\partial_t G}(s,t) \end{bmatrix}. \tag{3.53}$$

Comparing (3.52) and (3.53), we conclude

$$\hat{G}(s,t) = \hat{\beta}_0 \tag{3.54}$$

In our case, $\hat{\beta}_0 = \hat{\gamma}_{0,0}$ as has been shown above.

similarly, we can get,

$$h_G\widehat{\partial_s G}(s,t) = \hat{\gamma}_{1,0}. \tag{3.55}$$

Now, we have the equation,

$$\sup_{0\le s\le t\le 1} |\hat{G}_{(s,t)} - G_{(s,t)}|. \tag{3.56}$$

By replacing $\hat{G}_{(s,t)}$ by $(\hat{\gamma}_{(0,0)})$, the above equation becomes,

$$\sup_{0\le s\le t\le 1} |\hat{\gamma}_{(0,0)} - \gamma_{(0,0)}|. \tag{3.57}$$

Hence the proof. □

CHAPTER 4

Numerical Simulations

"There is no branch of mathematics, however abstract, which may not someday be applied to phenomena of the real world."

-Nikolaĭ Ivanovich Lobachevskiĭ

In this chapter we have provided some insights along with the graph into the numerical simulations run for the analysis of results obtained. Here we have discussed the simulation of n i.i.d. paths of diffusions $\{X_i\}_{i=1\cdots n}$ using Euler Maruyama numerical approximation technique having step size, $dt = 10^{-3}$. To do this we chose r random points $\{T_{ij}\}_{j=1\cdots r}$ in unit interval in increasing order for each X_i. The mathematical model that we have used for further calculation is

$$Y_{ij} = X_i(T_{ij}) + U_{ij} \quad i = 1 \cdots n, j = 1 \cdots r \tag{4.1}$$

where U_{ij} is i.i.d. Gaussian measurement with mean zero and finite variance ν^2. The performance was explored for values of n ranging from 100 to 1000, values of r ranging from 2 to 10, and for $\nu \in \{0, 0.05, 0.1\}$. We have used Epanechnikov kernel along with bandwidth $h = (n.r)^{(-1/5)}$ for both, surface and linear smoothing. For all (n, r, ν), 100 Monte Carlo simulations were performed, and the performance of estimators was checked by calculating the average square root of integrated square error obtained over these runs. This means that from every run we get $||\mu - \hat{\mu}||_{L^2([0,1])}$, $||\sigma^2 - \hat{\sigma}^2_T||_{L^2([0,1])}$, $||\sigma^2 - \hat{\sigma}^2_D||_{L^2([0,1])}$ and we evaluated the distribution followed by these errors over 100 iterations.

We discuss two examples of time-varying SDEs. The first one is a Brownian bridge where the drift is time-varying and the diffusion is constant. The second one is the time-inhomogeneous Ornstein Uhlenbeck process which has a drift that is sinusoidal time-varying and diffusion that is negative exponential time-varying. These two examples showcase the efficiency of the proposed estimators and also provide a comparison about how the performance changes based on the values chosen for n and r.

4.1 Brownian-Bridge

For this example, we have a Brownian-Bridge starting at 2 with time-varying drift i.e., $\mu(t) = \frac{-1}{1-t}$, and a constant diffusion where $\sigma = 1$. That is, we have,

$$\begin{cases} dX(t) = \frac{-1}{1-t} X(t)dt + dB(t), \ t \in [0, 1] \\ X(0) = 2. \end{cases} \tag{4.2}$$

As we can see, $\mu(t)$ is not well defined at t=1, hence the equation does not satisfy our assumptions. However, the proposed method can still deal with such cases as well which can be deduced from the first and second moments of the Brownian-Bridge. An important feature required for the local linear (surface) regression to fulfill the convergence rates in (3.3) is for the functions m and G to be smooth.

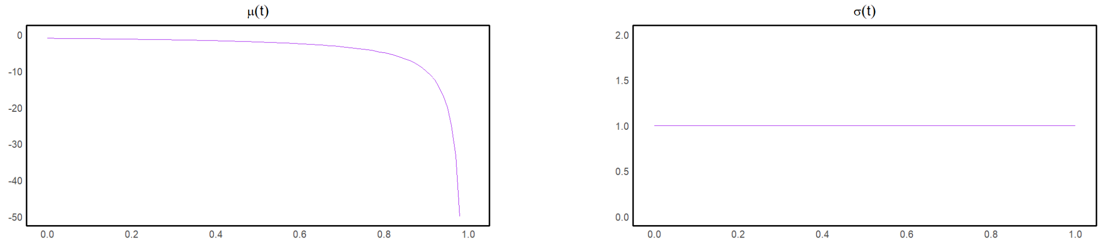

Figure 4.1: Drift and diffusion functions for Brownian Bridge process

The behavior of drift and diffusion estimators $\hat{\mu}$, $\hat{\sigma}_D$ and $\hat{\sigma}_T$ is studied across the number of points observed per curve, r taking the values 2, 3, 5 and 10 and number of curves, n taking the values 100, 200, 500 and 1000. Given below are the heatmap and boxplot for the average RISE of proposed estimators over 100 experiments for different values of n and r which are independent, with a variance of 0.05. In the heatmaps, the pink shades show a low average rise whereas the dark

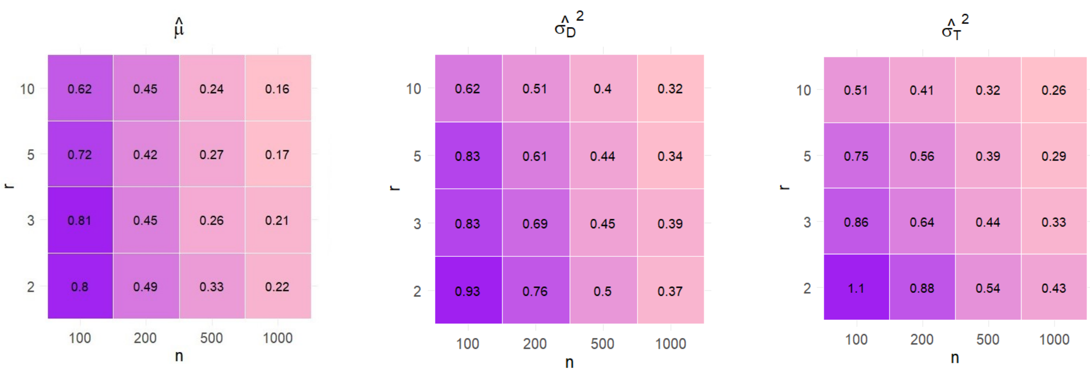

Figure 4.2: Heatmap illustration of average RISE for the estimators $\hat{\mu}(t)$, $\hat{\sigma_D}(t)$, and $\hat{\sigma_T}(t)$

purple shades show a high average rise. We can observe that for fixed values of r, increment in n gives more accurate estimates and as we increase r, we can notice much faster convergence rates. Even though the estimators' average RISE is given on the same scale, performance comparisons should be avoided. In fact, diffusion estimation necessitates knowledge of the covariance structure, whereas the nature of the functional object (a conditional average vs. a conditional variance) and the estimation process are fundamentally different.

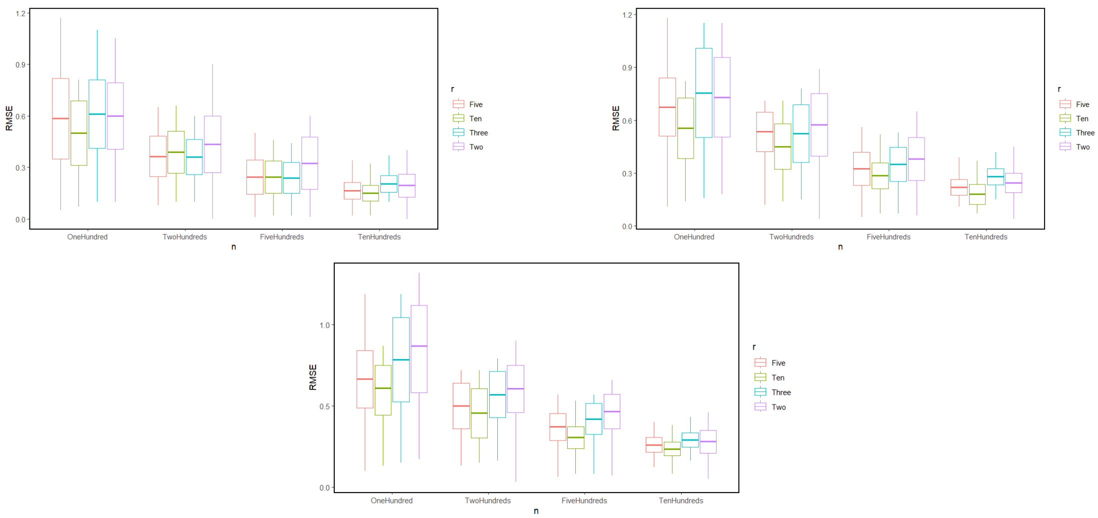

Figure 4.3: Boxplot of the average RISE for different values of (n,r)

4.2 Time-Inhomogeneous Ornstein Uhlenbeck

In this example, we have a time-varying SDE with sinusoidal drift and negative exponential diffusion,

$$\begin{cases} dX(t) = \frac{-1}{5}(1+\sin(2\pi t))X(t)dt + \sqrt{e^{(1-t^2)}}dB(t), \ t \in [0,1] \\ X(0) = 2, \end{cases} \tag{4.3}$$

to demonstrate the ability of this method to recover complex time-varying drift and diffusion coefficients. The rest of the discussions are the same as that of the above example.

The figures given thus are the heatmaps and boxplots for time-inhomogeneous Ornstein Uhlenbeck. For this case we have similar conclusions as that of the first example hence we have only provided the graphs.

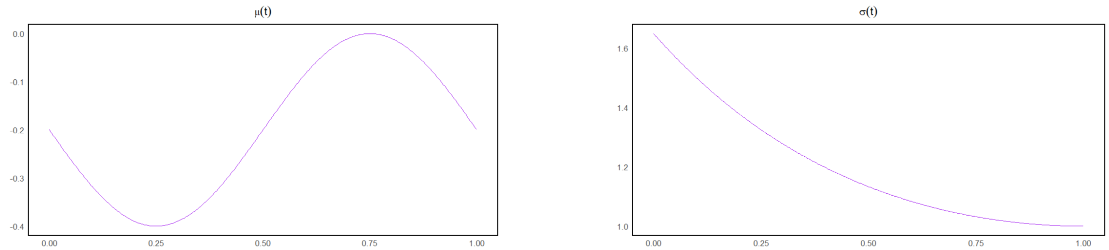

Figure 4.4: Drift and diffusion functions for Time-Inhomogeneous Ornstein Uhlenbeck

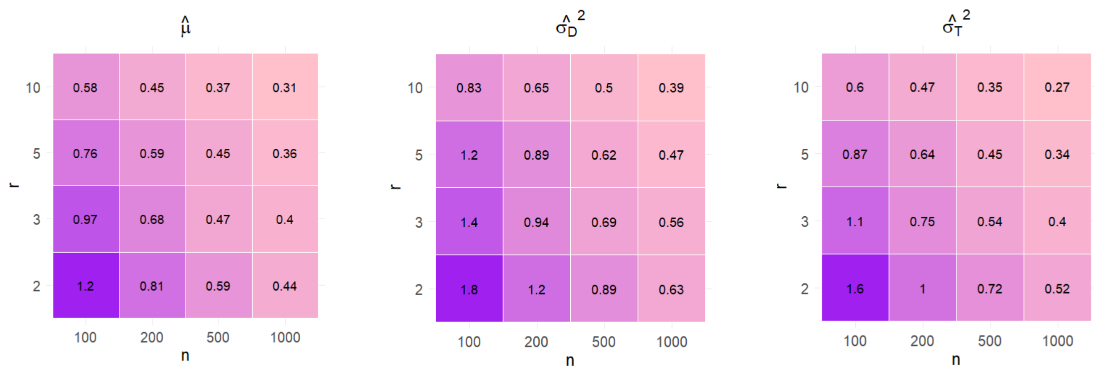

Figure 4.5: Heatmap illustration of average RISE for the estimators $\hat{\mu}(t)$, $\hat{\sigma_D}(t)$, and $\hat{\sigma_T}(t)$

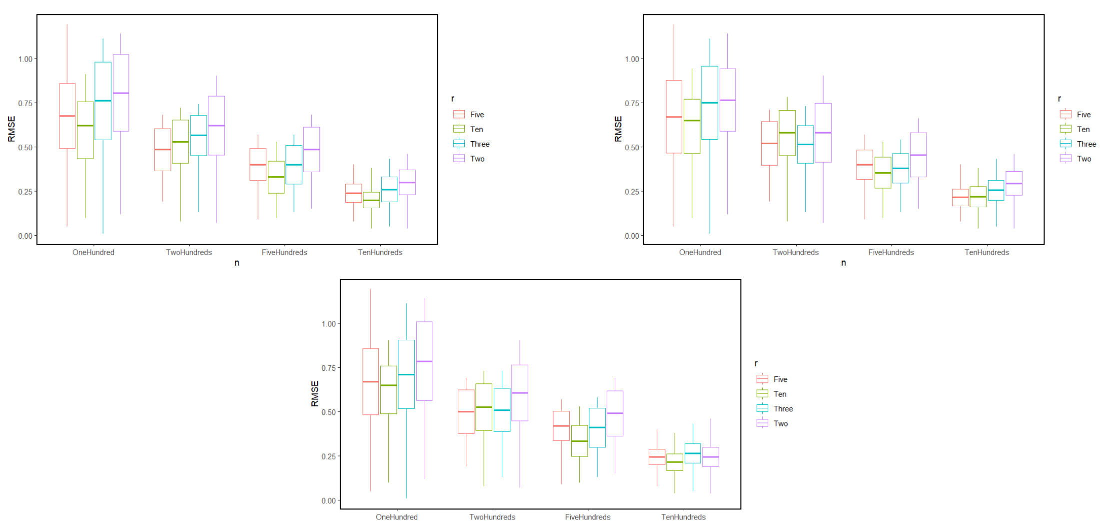

Figure 4.6: Boxplot of the average RISE for different values of (n,r)

CHAPTER 5

Conclusion and Future Scope

"The mathematician's patterns ... must be beautiful... beauty is the first test; there is no permanent place in the world for ugly mathematics".

-G. H. Hardy

In our study, we have worked with the stochastic differential equation of the form, $dX(t) = \mu(t)(X(t))^{\alpha}dt + \sigma(t)(X(t))^{\beta}dB(t)$ where $\alpha \in \{0, 1\}$ and $\beta \in \{0, 1/2, 1\}$. We have provided all the definitions and theorems required to understand the work presented. In this thesis, we have given the proof of theorem (3.3) in a much simpler manner in order to make it easier for the reader to grasp and follow along. Further, we have given two examples in which we estimated the drift and diffusion coefficients and put forth the RISE assessed with respect to the Monte Carlo simulations. These examples showed that the local polynomial estimators perform better when the values for n and r are increased and the RISE obtained for them is m=significantly lesser. For the estimators $\hat{\sigma_T}$ and $\hat{\sigma_D}$, some more analysis shows that σ_T gets a much lower RISE as it is tied to the complete covariance surface whereas, σ_D only takes the diagonal values.

The work given in this thesis can be extended to do the nonparametric estimation of the drift and diffusion coefficients of Stochastic Differential Equations (SDEs), subject to Lévy noise corruption. These areas are still open to researchers and offer a lot of scope for further research. A mathematical setting that anyone interested in taking this work further may consider is as follows.

Suppose that the Lévy noise is observed at discrete time points and that $\mathfrak{P} := \big(\Omega, \mathcal{F}, \mathbb{F} = \{\mathcal{F}\}_{t\geq 0}, P\big)$ is a filtered probability space satisfying the usual hypotheses. The following class of linear, time-dependent stochastic differential equations that are affected by additive Lévy noise can be studied.

$$dX(t) = \mu(t)X(t)\,dt + \sigma(t)\,dW(t) + \int_Z G(X(t-), z)\,\tilde{N}(dt, dz), \tag{5.1}$$

where $W(t)$ is an one-dimensional standard Brownian motion, $\mu(t)$ and $\sigma(t)$ are the drift and the diffusion coefficients defined on $\mathbb{R}$. Here, Z is a locally compact Polish space, and N is a Poisson random measure on $[0, \infty) \times Z$ with a σ-finite intensity measure $\lambda_\infty \otimes \nu$ on $\mathfrak{P}$. Moreover, λ_∞ is the Lebesgue measure on $[0, \infty)$, and ν is a σ-finite measure on Y, and $\tilde{N}([0, t] \times \mathcal{O})$ is the compensated Poisson random measure. The research on this is being conducted already but the area of study of SDEs with replications having noise corruption and estimation of drift, diffusion, and jump coefficients which are prone to Lévy noise are still wide open.

Bibliography

[1] On estimating the diffusion coefficient from discrete observations.

[2] Sharp adaptive drift estimation for ergodic diffusions: The multivariate case. *Stochastic Processes and their Applications*.

[3] Yacine Aït-Sahalia and Andrew W Lo. Nonparametric estimation of state-price densities implicit in financial asset prices. *The journal of finance*.

[4] G. Banon and Hung T. Nguyen. Recursive estimation in diffusion model. *SIAM Journal on Control and Optimization*.

[5] Carmen Broto and Esther Ruiz. Estimation methods for stochastic volatility models: a survey. *Journal of Economic surveys*.

[6] Zdzislaw Brzezniak and Tomasz Zastawniak. *Basic stochastic processes: a course through exercises*. Springer Science & Business Media, 2000.

[7] Fabienne Comte and Valentine Genon-Catalot. Drift estimation on non compact support for diffusion models. *Stochastic Processes and their Applications*.

[8] Jianqing Fan. *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*. Routledge, 2018.

[9] Hans-Georg Müller Fang Yao and Jane-Ling Wang. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*.

[10] Constantino A Garcia, Abraham Otero, Paulo Felix, Jesus Presedo, and David G Marquez. Nonparametric estimation of stochastic differential equations with sparse gaussian processes. *Physical Review E*.

[11] Marc Hoffmann. Adaptive estimation in diffusion processes. *Stochastic processes and their Applications*.

[12] Hui-Hsiung Kuo. *Stochastic differential equations*. Springer, 2006.

[13] Neda Mohammadi and Victor M Panaretos. Functional data analysis with rough sample paths? *Journal of Nonparametric Statistics*.

[14] Neda Mohammadi, Leonardo V. Santoro, and Victor M. Panaretos. Nonparametric estimation for SDE with sparsely sampled paths: an FDA perspective. *Stochastic Process. Appl.*

[15] Richard Nickl and Kolyan Ray. Nonparametric statistical inference for drift vector fields of multi-dimensional diffusions. 2020.

[16] Bernt Øksendal. *Stochastic differential equations*. Springer, 2003.

[17] Richard Stanton. A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk. *Journal of Finance*, 1997.

[18] Claudia Strauch. Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions. *Probability Theory and Related Fields*.

[19] Pham Dinh Tuan. Nonparametric estimation of the drift coefficient in the diffusion equation. *Series Statistics*.