# Consistency and Convergence of non-parametric estimation of drift and diffusion coefficients in SDEs from long stationary time-series

by
Xi Chen

A dissertation submitted to the Department of Mathematics,

College of Natural Sciences and Mathematics

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in Mathematics

Chair of Committee: Professor Ilya Timofeyev

Committee Member: Professor Andrew Török

Committee Member: Professor William Ott

Committee Member: Professor Rafail Abramov

University of Houston
May 2020

## ACKNOWLEDGMENTS

Time flies, it is time to say goodbye to my school year. I wish to thank all the people whose assistance was a milestone in the completion of this dissertation.

Firstly, I wish to express my sincere appreciation to my advisor Dr. Ilya Timofeyev. He led me to the field of applied math which needs a solid background in mathematical reasoning and computational skills. Moreover, he is a perfect mentor who gave me plenty of useful and constructive suggestions on work-life balance. Without his continual help, it would have been impossible for me to complete my dissertation and obtain the degree of Doctor of Philosophy.

Secondly, I would like to express my special regards to my committee members – Dr. William Ott (UH), Dr. Andrew Török (UH) and Dr. Rafail Abramov (UIC). I would like to thank them for their patience and guidance. Dr. William Ott impressed me with his fantastic course on theory of functions of a real variable and his interdisciplinary talk at the SIAM conference. As a student, I attended a course on stochastic differential equation given by Dr. Andrew Török and this course paved the way to my research area. Even though I did not take courses from Dr. Rafail Abramov, I would like to thank him for his time and feedback about my dissertation.

Thirdly, I would like to express my thanks to faculty, staff and friends. I would like to thank Dr. Ji who accepted me into the big family of the University of Houston. During my first school year here he shared his experiences and provided me with very useful advice. Dr. Pan whose earnest attitudes towards mathematics affected me deeply. Ms Katherine Vu, Dr. Meagan Carney, Ms Nhi (Amy) Kha and Ms Bennie Anderson who gave me lots of support. I wish to show my gratitude to Duong, Kayla and Daewa who held a baby shower for me and shared happiness together.

Last but not least, I wish to acknowledge the support and great love of my family, my husband, Peng Zhang; my parents and my daughter, Agnes. They kept me going on and this work would not have been possible without their inputs.

# ABSTRACT

We study the efficiency of non-parametric estimation of stochastic differential equations driven by Brownian motion (i.e., Diffusions) from long stationary trajectories. First, we introduce estimators based on conditional expectation which is motivated by the definition of Drift and Diffusion coefficients for SDEs. These estimators involve time- and space- discretization parameters for computing discrete analogs of expected values from discretely-sampled stationary data. The number of observational points is the third important computational parameter. Next, we derive bounds for the asymptotic behavior of $L^2$ errors for the Drift and Diffusion estimators. The asymptotic behavior is characterized when the number of observational points becomes infinite and observational time-step and bin size for spatial discretization of Drift and Diffusion coefficients tend to zero. Using our asymptotic analysis we are able to obtain practical guidelines for selecting computational parameters. Finally, we perform a series of numerical simulations which support our analytical investigation and illustrate practical guidelines for selecting near-optimal and computationally efficient values for computational parameters.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1 Introduction

The amount of observational and numerical data has been growing at an increasing rate during the last few decades. Now data demonstrates its vital position in our daily life. Not long ago we did not have a convenient and efficient ways to obtain information and obtaining quality observational and/or numerical data required a substantial amount of effort. Now, the amount of data is abundant. Presently it is very easy to develop and install all kinds of observational devices. Computing and data-storage capabilities also advanced significantly in the last few decades and nowadays computers are capable of producing enormous amounts of data. We are entering a data boosting epoch and data reflects many aspects of our daily life.

Here we address the problem of estimating Drift and Diffusion terms in stochastic differential equations driven by Brownian motion from observational data. This is a somewhat classical problem and there exists a considerable amount of literature on this topic. Here we concentrate on non-parametric estimation. The main motivation for our work is that until recent the past few decades there was a problem collecting enough data for successful application of this technique. Nowadays, with increasing amounts of observational and numerical data, some techniques should be revisited since they might become applicable in practice. We consider non-parametric estimators motivated by conditional expectations used in definitions of the Drift and Diffusion coefficients [7, 12]. For instance, non-parametric estimation of SDEs using similar ideas was applied to the one-dimensional problem in [4], but now the interest to this non-parametric purely data-driven technique has reemerged (e.g., [16–19, 31]) with potential applications to higher-dimensional models.

There are many practical problems where estimating an effective model from available observational or numerical data is of great interest. Such models can be used for future predictions as well as for the analysis of the underlying physics. The mathematical approach here applies primarily to temporal measurements. There are many possible areas of application in applied science and engineering such as econometrics, turbulence [20–22, 24, 32, 33], weather forecasting [23, 25, 38–44], climate studies [34–37], biology [16, 18, 31, 45, 46], reduced model on nonlinear dynamics [26–30], etc.

There are many branches related to the topic of my dissertation, such as time series analysis, differential equation, some topics in statistics, etc. Time series analysis [47–51] comprises methods for analyzing time series data which may have some internal structures, for instance, auto-correlation - the similarity between observations as a function of the time lag among them; seasonality - fluctuations that repeat regularly over time; and stationarity - time series where statistical properties do not change over time. The most closely related topic here is on autoregressive processes such as $AR(p)$ [59–63] or $GARCH$ [64–68].

There has been an effort to develop mathematical tools for effective learning of right-hand sides of differential equations from data. The main idea is quite similar to the overall goal of my dissertation - learn about dynamic rules which govern behavior of dynamic variables. Some recently developed approaches include Sparse Identification of Nonlinear Dynamics (SINDy) [69,70] and machine learning techniques [71].

Statistics [52–55] is one of the main scientific disciplines addressing various aspects of data science such as collection, management, illustration, analysis, interpretation, and presentation. Statistical inference [56] is one of the most popular areas, which includes estimation of parametric and non-parametric models. Parametric estimation is widely a used method based on the hypothesis of fitting empirical data into a known distribution with hidden properties or generic structures. The topic of my dissertation is related to non-parametric estimation. Non-parametric estimation is a statistical method that allows the functional form of a fit to data to be obtained in the absence of any guidance or constraints from theory [15] or any underlying assumptions. The non-parametric approach is more flexible and compatible with any data sets. However, this approach is sometimes more difficult to validate and it typically requires much more data compared to parametric techniques [74–78].

Finally, regression analysis is a subset of statistical modeling which deals with estimating the functional form of relationships between dependent and independent variables. It is widely used in finance, investing, and other disciplines that attempts to determine the importance of different variables for generating predictions. Generally, regression can be separated into two groups - (i) linear regression models (e.g., [53, 57, 58]) which was particular popular several decades ago when computers were non-existent and (ii) nonlinear regression models (e.g., [72, 73]) which can tackle more general problems.

Every estimator has internal computational parameters which control the quality of estimation. Of the most obvious parameters is the number of points, $M$, available for evaluating the estimator. Another parameter which is common is the observation time-step for collecting the discrete data, $\Delta t$. In our application of non-parametric estimation there is a third parameter, $\Delta x$, related to space-discretization for estimating the Drift and Diffusion coefficients. One of the most important practical problems is the optimal choice of these internal computational parameters such that it minimizes the computational cost while maximizing the efficiency of estimators. In my dissertation I use analytical estimates derived in the limiting regime of $M \to \infty$, $\Delta t$, $\Delta x \to 0$ to analyze the optimal selection of these computational parameter. It turns out that accuracy of the Drift and Diffusion estimators depends on these three parameters in a non-trivial way. Thus, we develop practical guidelines for selecting these computational parameters.

The rest of the dissertation is organized as follows - in Chapter 2 background material from real analysis and stochastic differential equations is introduced first, then estimators for the Drift and Diffusion terms are introduced, and properties of the truncated density are discussed in more detail. Bias of these estimators is also discussed in detail. The most important important analytical contribution is the analysis of the Mean-Square Error (MSE) for the Drift and Diffusion estimators in sections 2.9 and 2.10, respectively. Numerical simulations are presented in Chapter 3 where we analyze the behavior of Absolute Error and Mean-Square-Error (MSE) for two particular models. Our simulations support our analytical conclusions and provide practical guidelines for the optimal choice of computational parameters $M$, $\Delta t$, and $\Delta x$ for the Drift and Diffusion estimators. Also, we use several regression techniques to estimate the functional form of the Drift and Diffusion terms from our non-parametric estimators. In particular, we compare and contrast polynomial fit (least square [53]) and two shrinkage methods - Lasso [53, 57] and Ridge regression [53, 58] in the Root-Mean-Square-Error (RMSE) sense.

## 2 Theoretical formulation

Our goal is to derive practical formulas for computing estimators of the Drift and Diffusion coefficients in SDEs from discretely sampled stationary time-series. To this end, we consider one-dimensional Diffusions (stochastic differential equations driven by Brownian motion) and use the definition of Drift and Diffusion coefficients to define our estimators. These estimators can be easily generalized to higher dimensions, but the analysis becomes much more difficult. Thus, we stay in the context of one-dimensional models.

We introduce our estimators in (10) and (11); the main theoretical results contain an analysis of Bias (sections 2.6 and 2.7) and Mean-Square-Error (sections 2.9 and 2.10). The analysis of the MSE for the Drift and Diffusion estimators is much more challenging since it requires estimating the behavior of many correlated terms and products of stochastic integrals. Results for the asymptotic behavior of the MSE are summarized in sections 2.9.1 and 2.10.1.

SDEs are widely used in the fields of finance, physics, meteorology, mathematical biology, etc. With the increase of observational data in recent years, estimation of effective processes (including SDEs) has become an important practical task. In this chapter we lay the main theoretical results which have practical consequences on how to determine computational parameters used in our Drift and Diffusion estimators.

### 2.1 Known inequalities

In this section we list some well-known inequalities for future reference.

**Inequality 2.1** (Hölder inequality in 6.6.2 of [5]). *Suppose* $1 < p < \infty$ *and* $p^{-1} + q^{-1} = 1$ *(that is,* $q = p/(p-1)$*). If* $f$ *and* $g$ *are measurable functions on arbitrary set* $X$*, then*

$$\|fg\|_1 \leq \|f\|_p \|g\|_q. \tag{1}$$

*In particular, if* $f \in L^p$ *and* $g \in L^q$*, then* $fg \in L^1$*, and in this case equality holds in (1) iff* $\alpha|f|^p = \beta|g|^q$ *a.e. for some constants* $\alpha$*,* $\beta$ *with* $\alpha\beta \neq 0$.

Specifically, take $p = q = 2$, we have

**Inequality 2.2** (Cauchy-Schwarz Inequality). *If $f$ and $g$ are measurable functions on an arbitrary set $X$, then*

$$\|fg\|_1 \leq \|f\|_2 \|g\|_2. \tag{2}$$

**Inequality 2.3** (Minkowski's Inequality in 6.6.5 of [5]). *If $1 \leq p \leq \infty$ and $f, g \in L^p$, then*

$$\|f + g\|_p \leq \|f\|_p + \|g\|_p. \tag{3}$$

## 2.2 Preliminary results for SDEs

In this section we recall some background theoretical results for SDEs. Consider a one-dimensional SDE driven by Brownian motion

$$dX_t = A(X_t)dt + B(X_t)dW_t. \tag{4}$$

Here $A(X_t)$ and $B(X_t)$ are called the Drift and Diffusion coefficients, respectively. SDEs driven by Brownian motion are often called Diffusion processes or simply Diffusions. The equation above is understood in an integral sense, i.e.,

$$X(t) = X_0 + \int_0^t A(X(s))\, ds + \int_0^t B(X(s))\, dW_s.$$

Here we treat all integrals with respect to the Brownian motion (a.k.a. Wiener process) in the Itô sense [1, 7, 12].

We would like to recall the following basic properties for Diffusions -

1. Functions $A$ and $B$ have to be Lipschitz to ensure existence and uniqueness of solution of

the SDE in (4) [1, 12]

$$\begin{aligned} |A(x) - A(y)| &\leq K_A|x - y|, \\ |B(x) - B(y)| &\leq K_B|x - y|, \end{aligned}$$

where $K_A$ and $K_B$ are some constants.

2. The Drift and Diffusion coefficients can be defined using conditional expectations of the process $X_t$ [7–11]

$$A(x) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{E}\left[ X_{t+\Delta t} - x \middle| X_t = x \right], \tag{5}$$

$$B(x)^2 = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{E}\left[ (X_{t+\Delta t} - x)^2 \middle| X_t = x \right]. \tag{6}$$

3. We can also easily derive the Fokker-Planck equation [2, 3, 7] for the conditional probability $f(x,t) = p(x,t|x_0,t_0)$ for any initial $x_0$, $t_0$.

$$\frac{\partial f(x,t)}{\partial t} = -\frac{\partial}{\partial x}[A(x)f(x,t)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x)f(x,t)] \tag{7}$$

with the initial condition $p(x,t_0|x_0,t_0) = \delta(x - x_0)$. The transition probability density is defined as

$$Pr\{X(t) \in A | X(0) = x_0\} = \int_A p(x,t|x_0,t_0)\,dx.$$

Equations (4) and (7) are regarded as complementary to each other.

### 2.3 Estimators

In this section we introduce estimators for the Drift and Diffusion coefficients. These estimators are based on definitions (5) and (6), but take into account the fact that we use discrete observations. In particular, we cannot implement conditioning in (5) and (6) directly for discrete data. Therefore, we introduce bins to compute discrete versions of conditional expectation in (5) and (6). We assume that all bins have the same width and we denote it as $\Delta x$. Our approach can be easily generalized to bins of varying size, but this does not affect conclusions reached here because our analysis is carried out separately for each bin. We introduce a discrete mesh $x_k$, $k = 1, \ldots, K$ with $x_{k+1} - x_k = \Delta x$. Points $x_k$ represent centers of bins $[x_k - \Delta x/2, x_k + \Delta x/2]$ which are used for conditioning when computing discrete analogs of expected values.

We also assume that data is sampled with time-step $\Delta t$ and a data set contains exactly $M$ observations for each bin. Therefore the total number of observational time-instances is $M \times K$ where $K$ is the number of bins defined above.

We define discrete estimators analogously to expressions in (5) and (6)

$$\begin{aligned}
\hat{A}(x_k) &= \frac{1}{\Delta t}\frac{1}{M}\sum_{j=1}^{MK}(X_{t_j+\Delta t} - X_{t_j})\mathbb{1}(X_{t_j}, k), \\
\hat{B}^2(x_k) &= \frac{1}{\Delta t}\frac{1}{M}\sum_{j=1}^{MK}\left(X_{t_j+\Delta t} - X_{t_j}\right)^2 \mathbb{1}(X_{t_j}, k),
\end{aligned}$$

where the indicator function is defined as

$$\mathbb{1}(X_t, k) = \begin{cases} 1, & X_t \in Bin_k, \\ 0, & X_t \notin Bin_k, \end{cases} \tag{8}$$

and

$$Bin_k \simeq \left[x_k - \frac{\Delta x}{2}, x_k + \frac{\Delta x}{2}\right]. \tag{9}$$

The estimators above can be rewritten slightly differently as

$$\hat{A}(x_k) = \frac{1}{\Delta t}\frac{1}{M}\sum_{j \in M_k}(X_{t_j+\Delta t} - X_{t_j}), \tag{10}$$

$$\hat{B}^2(x_k) = \frac{1}{\Delta t}\frac{1}{M}\sum_{j \in M_k}\left(X_{t_j+\Delta t} - X_{t_j}\right)^2, \tag{11}$$

where the set $M_k = \{j : \mathbb{1}(X_{t_j}, k) = 1\}$, and $card(M_k) = M$. Set $M_k$ is a set of indexes such that

$X_{t_j} \in Bin_k$ and contains exactly $M$ time-instances. The indicator function $\mathbb{1}(X_{t_j}, k)$ is defined in (8).

Here, the indicator function $\mathbb{1}(X_t, k)$ plays the role of conditioning in expressions (10), (11), but the conditioning is done on the interval $Bin_k$ instead of a particular value. We also impose $card(M_k) = M$ for all $k$, which means that we consider the situation when the number of time-instances for estimating the Drift and Diffusion coefficients does not depend on $x_k$. This implies that for all bins data always contains at least $M$ time-instances $t_j$ such that $X_{t_j} \in Bin_k$ for all $k$. In practice, such a situation is likely to occur when none of the $x_k$ are in the tails of stationary distribution $\rho(x)$, e.g., $\max(x_k) - \min(x_k) \approx stddev(\rho(x))$. This is exactly the situation for many practical applications when observational data is produced by numerical simulations in stationary regime or long empirical observations of stationary processes and rare events are unlikely to be a part of the observed trajectory. Stationary turbulence is one example of such applications. Another application are the long-term climatology observations and simulations/observation of macro molecular dynamics (e.g., protein folding).

### 2.4 Truncated density

We assume that the SDE (4) admits a unique stationary distribution $\rho(x)$ which is the solution of the Fokker-Planck equation

$$0 = -\frac{\partial}{\partial x}[A(x)\rho(x)] + \frac{1}{2}\frac{\partial^2}{\partial x^2}[B(x)\rho(x)].$$

We define a restricted density corresponding to $Bin_k$ as

$$p_k(x) = \begin{cases} c\rho(x), & \text{if } x \in Bin_k, \\ 0, & \text{if } x \notin Bin_k, \end{cases} \quad \text{where } \ c = \left( \int\limits_{x_k - \frac{\Delta x}{2}}^{x_k + \frac{\Delta x}{2}} \rho(s) ds \right)^{-1}. \tag{12}$$

The corresponding cumulative distribution function is

$$P_k(x) = \int_{x_k - \frac{\Delta x}{2}}^{x} p_k(s)ds.$$

Since $p_k(x)$ is the density of the truncated distribution, we have

$$P_k\left(x_k - \frac{\Delta x}{2}\right) = 0, \quad P_k\left(x_k + \frac{\Delta x}{2}\right) = 1.$$

**Lemma 2.4.** *Assuming that $\rho(x)$ is a sufficiently differentiable function, then the constant $c$ in the definition of the truncated density* (12) *is given to the leading order by*

$$c = \frac{1}{\rho(x_k)\Delta x} + O(\Delta x).$$

*Proof.* Expanding $\rho(s)$ at $x_k$ using Taylor series, we obtain

$$\rho(s) = \rho(x_k) + \rho'(x_k)(s - x_k) + \frac{\rho''(x_k)}{2}(s - x_k)^2 + \cdots .$$

Then

$$\begin{aligned} c^{-1} = \int_{x_k - \frac{\Delta x}{2}}^{x_k + \frac{\Delta x}{2}} \rho(s)ds &= \int_{x_k - \frac{\Delta x}{2}}^{x_k + \frac{\Delta x}{2}} \rho(x_k)ds + \rho'(x_k) \int_{x_k - \frac{\Delta x}{2}}^{x_k + \frac{\Delta x}{2}} s - x_k ds + \frac{\rho''(x_k)}{2} \int_{x_k - \frac{\Delta x}{2}}^{x_k + \frac{\Delta x}{2}} (s - x_k)^2 ds + \cdots \\ &= \rho(x_k)\Delta x + \frac{\rho''(x_k)}{24}(\Delta x)^3 + O\left((\Delta x)^5\right) \\ &= \Delta x \left(\rho(x_k) + \frac{\rho''(x_k)}{24}(\Delta x)^2 + O\left((\Delta x)^4\right)\right). \end{aligned}$$

If we use

$$\frac{1}{x + \varepsilon} \approx \frac{1}{x} - \frac{\varepsilon}{x^2} \quad \text{for } \varepsilon << 1$$

then we obtain

$$\begin{aligned} c &= \frac{1}{\Delta x}\left[\frac{1}{\rho(x_k)} - \frac{1}{\rho^2(x_k)}\left(\frac{\rho''(x_k)}{24}(\Delta x)^2 + O\left((\Delta x)^4\right)\right)\right] \\ &= \frac{1}{\Delta x}\left[\frac{1}{\rho(x_k)} - \frac{1}{\rho^2(x_k)}\frac{\rho''(x_k)}{24}(\Delta x)^2 + O\left((\Delta x)^4\right)\right]. \end{aligned}$$

□

Rearranging the expansion result for the density $\rho$, we obtain

$$\rho(x_k)\Delta x = 1 - \frac{\rho''(x_k)}{24}(\Delta x)^3 + O\left((\Delta x)^5\right). \tag{13}$$

### 2.4.1 Moments for the truncated density function

In this section, we analyze moments of the truncated density. In particular, we formulate the following lemma

**Lemma 2.5.** *For any function $f(x) \in C^2\left[x_k - \frac{\Delta x}{2}, x_k + \frac{\Delta x}{2}\right]$, the conditional expectation of $f(x)$ is given by*

$$\mathbb{E}[f(X_t)|X_t \in Bin_k] = f(x_k) + C\,(\Delta x)^2 + O((\Delta x)^4), \tag{14}$$

*where $X_t$ is the stationary solution of* (4) *and*

$$C = \frac{f''(x_k)\rho(x_k) + 2f'(x_k)\rho'(x_k)}{24\rho(x_k)}.$$

*Proof.* From the definition of the conditional expectation in (14)

$$
\begin{aligned}
\mathbb{E}[f(X_t)|X_t \in Bin_k] = \int_{-\infty}^{\infty} f(x)p_k(x)\,dx = c \int_{x_k-\Delta x/2}^{x_k+\Delta x/2} f(x)\rho(x)\,dx \\
= \quad c\left[f(x_k)\rho(x_k)\Delta x + \frac{f''(x_k)\rho(x_k) + 2f'(x_k)\rho'(x_k) + f(x_k)\rho''(x_k)}{24}(\Delta x)^3 + O((\Delta x)^5)\right] \\
= \quad f(x_k) + \frac{f''(x_k)\rho(x_k) + 2f'(x_k)\rho'(x_k)}{24\rho(x_k)}(\Delta x)^2 + O((\Delta x)^4),
\end{aligned}
$$

where we used Lemma 2.4. □

**Low-order moments.** From Lemma 2.5, we have the following expression for the first four moments

$$
\begin{aligned}
\mathbb{E}[X_t|X_t \in Bin_k] &= x_k + \frac{\rho'(x_k)(\Delta x)^2}{12\rho(x_k)} + O((\Delta x)^4), \\
\mathbb{E}[X_t^2|X_t \in Bin_k] &= x_k^2 + \frac{\rho(x_k) + 2f'(x_k)\rho'(x_k)}{12\rho(x_k)}(\Delta x)^2 + O((\Delta x)^4), \\
\mathbb{E}[X_t^3|X_t \in Bin_k] &= x_k^3 + \frac{x_k(\rho(x_k) + x_k^2\rho'(x_k))}{4\rho(x_k)}(\Delta x)^2 + O((\Delta x)^4), \\
\mathbb{E}[X_t^4|X_t \in Bin_k] &= x_k^4 + \frac{x_k^2(3\rho(x_k) + 2x_k\rho'(x_k))}{6\rho(x_k)}(\Delta x)^2 + O((\Delta x)^4).
\end{aligned}
$$

If we assume that $A(x)$ and $B(x)$ are twice differentiable, then using Lemma 2.5 we also straightforwardly compute the expected values of the **Drift term**

$$
\begin{aligned}
\mathbb{E}_k[A(X_t)] = \mathbb{E}[A(X_t)|X_t \in Bin_k] = c \int_{x_k-\frac{\Delta x}{2}}^{x_k+\frac{\Delta x}{2}} A(s)\rho(s)ds \\
= \quad A(x_k) + \frac{2A'(x_k\rho'(x_k)) + A''(x_k)\rho(x_k)}{24\rho(x_k)}(\Delta x)^2 + O\left((\Delta x)^4\right)
\end{aligned}
$$

and the **Diffusion term**

$$\mathbb{E}_k[B^2(X_t)] = \mathbb{E}\left[B^2(X_t)|X_t \in Bin_k\right] = c \int_{x_k-\frac{\Delta x}{2}}^{x_k+\frac{\Delta x}{2}} B^2(s)\rho(s)ds$$

$$= \quad B^2(x_k) + \frac{2B(x_k)B'(x_k)\rho'(x_k) + [(B'(x_k))^2 + B(x_k)B''(x_k)]\rho(x_k)}{12\rho(x_k)}(\Delta x)^2 + O\left((\Delta x)^4\right).$$

### 2.4.2 Behavior of truncated density for small bin size

In this section, we study behavior of truncated density $p_k$ in (12) for small bin size. In particular, the truncated density approximates the delta function $\delta(x - x_k)$ for small $\Delta x$ (see Lemma 2.5). Thus, we want to understand the rate of growth (with respect to $\Delta x$ ) for the conditional density $p_k(x)$ as $\Delta x \to 0$.

Recall from Lemma 2.4

$$c = \left(\rho(x_k)\Delta x + \frac{\rho''(x_k)}{24}(\Delta x)^3 + O\left((\Delta x)^5\right)\right)^{-1}.$$

Next, expanding $\rho(x)$ at $x_k$ using Taylor series, we obtain

$$\rho(x) = \rho(x_k) + \rho'(x_k)(x - x_k) + \frac{\rho''(x_k)}{2}(x - x_k)^2 + \cdots .$$

Recall that $p(x) = c\rho(x)$ for $x \in Bin_k$ and combining the two expressions above we obtain

$$p_k(x) \approx \frac{\rho(x_k)}{\Delta x}\left(\rho(x_k) + \frac{\rho''(x_k)}{24}(\Delta x)^2 + O\left((\Delta x)^4\right)\right)^{-1}.$$

Similarly,

$$\begin{aligned} p_k'(x) &\approx \frac{\rho'(x_k)}{\Delta x}\left(\rho(x_k) + \frac{\rho''(x_k)}{24}(\Delta x)^2 + O\left((\Delta x)^4\right)\right)^{-1} \\ p_k''(x) &\approx \frac{\rho''(x_k)}{\Delta x}\left(\rho(x_k) + \frac{\rho''(x_k)}{24}(\Delta x)^2 + O\left((\Delta x)^4\right)\right)^{-1} \end{aligned}$$

Suppose that $\rho(x_k) \neq 0$, then we get

$$\frac{1}{1+\frac{\rho''(x_k)}{24\rho(x_k)}(\Delta x)^2+\frac{O\big((\Delta x)^4\big)}{\rho(x_k)}} \longrightarrow 1 \quad \text{as} \quad \Delta x \to 0$$

and we can formally state the following theorem

**Theorem 2.6.** *Suppose $\rho(x_k) \neq 0$ and $\rho(x_k) \in C^2\left[x_k - \frac{\Delta x}{2}, x_k + \frac{\Delta x}{2}\right]$ then*

$$\begin{aligned} p_k(x)\Delta x &\longrightarrow C_1, \\ p_k'(x)\Delta x &\longrightarrow C_2, \\ p_k''(x)\Delta x &\longrightarrow C_3 \end{aligned}$$

*as $\Delta x \to 0$, where $C_1$, $C_2$ and $C_3$ are finite constants.*

### 2.4.3 Comment about the number of points $M$ for computing Drift and Diffusion estimators for each bin

In this work we treat the number of time-instances which fall into each bin, $M$, as constant. In practice, one is often given a fixed dataset which contains a long stationary trajectory with a fixed total number of discrete time-instances for computing estimators. In such a situation there is little flexibility in choosing how many points falls in each bin and in such cases $M(x_k) \equiv card(M_k)$ should be treated as random. This implies that the summations in (10) and (11) become random sums. Therefore, theoretically, all formulas for the bias and $L^2$ errors should contain expectations with respect to the distribution for $M_k$. One can compute the mean of $M(x_k)$ as

$$\mathbb{E}[M(x_k)/M_{total}] = \int_{x_k-\Delta x/2}^{x_k+\Delta x/2} \rho(x)dx,$$

where $\rho(x)$ is the stationary distribution of the process, and $M_{total}$ (fixed and not random) is the total number of time-instances in the stationary trajectory. By Jensen's inequality ($\phi(\mathbb{E}[X]) \leq \mathbb{E}[\phi(x)]$ for any convex $\phi$)

$$\mathbb{E}\left[\frac{1}{M(x_k)}\right] \geq \frac{C(x_k)}{M_{total}},$$

where $C(x_k)$ depends on bin $k$ and stationary density $\rho(x)$. Expectations with respect to random $M(x_k)$ can be simplified using conditional expectations, e.g.

$$\begin{aligned}
\mathbb{E}\hat{A}(x_k) &= \mathbb{E}\left[\frac{1}{\Delta t}\frac{1}{M(x_k)}\sum_{j\in M_k}(X(t_j+\Delta t)-X(t_j))\mathbb{1}(X_{t_j},k)\right] \\
&= \mathbb{E}_{M(x_k)}\left[\mathbb{E}\left[\frac{1}{\Delta t}\frac{1}{M(x_k)}\sum_{j\in M_k}(X(t_j+\Delta t)-X(t_j))\mathbb{1}(X_{t_j},k)\Big| M(x_k)=M\right]\right] \\
&= \mathbb{E}_{M(x_k)}\left[\frac{1}{\Delta t}\frac{1}{M}\sum_{j=1}^{M}\mathbb{E}\left[(X(t_j+\Delta t)-X(t_j))\mathbb{1}(X_{t_j},k)\right]\right] \\
&= \frac{1}{\Delta t}\mathbb{E}\left[(X(t+\Delta t)-X(t))\mathbb{1}(X_t,k)\right],
\end{aligned}$$

where we reordered points in the data set to make the summation explicit. If $M(x_k)$ is treated as random, the analysis of the asymptotic properties of the Mean Squared Error (MSE) for our estimators becomes rather difficult since leading-order terms in those expressions involve $M^{-1}(x_k)$. Using formulas for the conditional expectation we can simplify calculation of the total MSE using condition $M(x_k) = m$ and then take the expectation over $M(x_k)$. However, computing the expectation of $M^{-1}(x_k)$ depends on the particular form of the stationary distribution $\rho(x)$. In addition, if the total number of points in the stationary trajectory is large, we expect that $M(x_k)$ would not deviate significantly from the mean and expressions for a fixed $M$ provide sufficient insight into behavior of our estimators.

## 2.5 Truncated Itô-Taylor expansion

In this section we introduce the main technical tool for the analysis of our estimators. In particular, we need to analyze various moments of $(X_{t_j+\Delta t} - X_{t_j})\mathbb{1}(X_{t_j}, k)$. One can proceed by analyzing the joint density of $X_{t_j}$ and $X_{t_j+\Delta t}$. This requires an explicit knowledge (either exact or approximate) of the transition probability density which is equivalent to solving the Fokker-Planck equation for short times, $\Delta t$.

Here we follow an alternative approach and introduce stochastic Itô-Taylor expansions for $X_{t_j+\Delta t}$. This allows us to simplify analytical computations because we no longer need the joint density of $X_{t_j}$ and $X_{t_j+\Delta t}$ to carry out the analysis. Instead, moments of $(X_{t_j+\Delta t} - X_{t_j})\mathbb{1}(X_{t_j}, k)$ can be estimated (up to certain order in $\Delta t$) from the stationary density of $X_{t_j}$ and properties of the Brownian motion on the interval $[t_j, t_j + \Delta t]$.

The Itô-Taylor expansion of $X_{t_j+\Delta t}$ with respect to stochastic integrals is defined as (see [14])

$$\begin{aligned}
X_{t_j+\Delta t} &= X_{t_j} + A_0(X_{t_j})I_{(0),j} + A_1(X_{t_j})I_{(1),j} + A_2(X_{t_j})I_{(0,0),j} + A_3(X_{t_j})I_{(0,1),j} + \\
&\quad A_4(X_{t_j})I_{(1,0),j} + A_5(X_{t_j})I_{(1,1),j} + A_6(X_{t_j})I_{(1,1,1),j} + \\
&\quad A_7(X_{t_j})I_{(0,1,1),j} + A_8(X_{t_j})I_{(1,0,1),j} + A_9(X_{t_j})I_{(1,1,0),j} + \dots
\end{aligned}$$

where $I_{(\cdot),j}$ are multiple stochastic integrals and coefficients $A_k(X_{t_j})$ are computed explicitly using the Drift and Diffusion coefficients and their derivatives.

For our analysis we can truncate the Itô-Taylor expansion at a certain order. The motivation for truncating higher-order terms will be given later in this section. Thus, we consider the truncated Itô-Taylor expansion of the following form

$$\begin{aligned}
X_{t_j+\Delta t} &\approx X_{t_j} + A_0(X_{t_j})I_{(0),j} + A_1(X_{t_j})I_{(1),j} + A_2(X_{t_j})I_{(0,0),j} + A_3(X_{t_j})I_{(0,1),j} \\
&\quad + A_4(X_{t_j})I_{(1,0),j} + A_5(X_{t_j})I_{(1,1),j} + A_6(X_{t_j})I_{(1,1,1),j} \\
&= X_{t_j} + \sum_{q=0}^{6} A_q(X_{t_j})I_{\alpha_q,j}, \qquad (15)
\end{aligned}$$

where $\alpha_q$ is the index used to define the corresponding stochastic integral (i.e., $\alpha_0 = (0)$, $\alpha_1 = (1)$, $\alpha_2 = (0,0)$, etc.) and $A_0(x) \equiv A(x)$ and $A_1(x) \equiv B(x)$ and other terms in the expansion are given by

$$\begin{aligned}
A_2(X_{t_j}) &= A(X_{t_j})A'(X_{t_j}) + \frac{1}{2}B^2(X_{t_j})A''(X_{t_j}),\\
A_3(X_{t_j}) &= A(X_{t_j})B'(X_{t_j}) + \frac{1}{2}B^2(X_{t_j})B''(X_{t_j}),\\
A_4(X_{t_j}) &= B(X_{t_j})A'(X_{t_j}), \qquad A_5(X_{t_j}) = B(X_{t_j})B'(X_{t_j}),\\
A_6(X_{t_j}) &= B(X_{t_j})\left((B'(X_{t_j}) + B(X_{t_j})B''(X_{t_j})\right).
\end{aligned}$$

Multiple stochastic integrals are defined as

$$\begin{aligned}
I_{(0),j} &= \int_t^{t+\Delta t} dt', \qquad I_{(1),j} = \int_t^{t+\Delta t} dW(t'),\\
I_{(0,0),j} &= \int_t^{t+\Delta t}\int_t^s dt' ds, \qquad I_{(0,1),j} = \int_t^{t+\Delta t}\int_t^s dt' dW,\\
I_{(1,0),j} &= \int_t^{t+\Delta t}\int_t^s dW(t')ds, \qquad I_{(1,1),j} = \int_t^{t+\Delta t}\int_t^s dW(t')dW(s),\\
I_{(1,1,1),j} &= \int_t^{t+\Delta t}\int_t^s\int_t^{t'} dW(r)dW(t')dW(s).
\end{aligned}$$

We can easily see that integrals $I_{(0),j}$ and $I_{(0,0),j}$ are deterministic and are given by $I_{(0),j} = \Delta t$, $I_{(0,0),j} = \Delta t^2/2$. Other integrals are stochastic, but some of them can also be computed explicitly in terms of $\Delta W_{j+1} = W_{t_j+\Delta t} - W_{t_j}$. For instance, $I_{(1),j} = \Delta W_{j+1}$. We can also compute $I_{(1,1),j}$ using a simple lemma below.

**Lemma 2.7.** *The stochastic integral $I_{(1,1),j}$ can be computed explicitly as*

$$I_{(1,1),j} = \frac{1}{2}\left((\Delta W_{j+1})^2 - \Delta t\right),$$

*where*

$$I_{(1,1),j} = \int_{t_j}^{t_j+\Delta t}\int_{t_j}^s dW_{t'} dW_s \quad \text{and} \quad \Delta W_{j+1} = W_{t_j+\Delta t} - W_{t_j}.$$

*Proof.* Using Integration by parts and the properties of stochastic integral, we have

$$\begin{aligned}
I_{(1,1),j} &= \int_{t_j}^{t_j+\Delta t}\int_{t_j}^{s} dW_{t'}dW_s = \int_{t_j}^{t_j+\Delta t}\Big[W_s - W_{t_j}\Big]\,dW_s \\
&= \int_{t_j}^{t_j+\Delta t} W_s dW_s - W_{t_j}\Delta W_{j+1} \\
&= \frac{1}{2}\int_{t_j}^{t_j+\Delta t}\Big[dW_s^2 - dt\Big] - W_{t_j}\Delta W_{j+1} \\
&= \frac{1}{2}\Big[W_{t_j+\Delta t}^2 - W_{t_j}^2 - \big(t_j + \Delta t - t_j\big)\Big] - W_{t_j}\Big(W_{t_j+\Delta t} - W_{t_j}\Big) \\
&= \frac{1}{2}\Big[W_{t_j+\Delta t} - W_{t_j}\Big]^2 - \frac{1}{2}\Delta t \\
&= \frac{1}{2}\Big[(\Delta W_{j+1})^2 - \Delta t\Big].
\end{aligned}$$

□

Here we also list some of the properties stochastic integrals

$$\begin{aligned}
& I_{(0),j} = \Delta t, \quad I_{(0,0),j} = \frac{\Delta t^2}{2}, \\
& \mathbb{E}I_{(1),j} = \mathbb{E}I_{(1,0),j} = \mathbb{E}I_{(0,1),j} = \mathbb{E}I_{(1,1,1),j} = \mathbb{E}I_{(1,1),j} = 0, \\
& \mathbb{E}I_{(1),j}^2 = \Delta t, \qquad \mathbb{E}I_{(1,1),j}^2 = \frac{(\Delta t)^2}{2}, \qquad \mathbb{E}I_{(1,1,1),j}^2 = O(\Delta t^3) \\
& \mathbb{E}I_{(0,1),j}^2 = \mathbb{E}I_{(1,0),j}^2 = \frac{(\Delta t)^3}{3}, \qquad \mathbb{E}\Big[I_{(0,1),j}I_{(1,0),j}\Big] = O(\Delta t^3), \\
& \mathbb{E}\Big[I_{(1),j}I_{(0,1),j}\Big] = \mathbb{E}\Big[I_{(1),j}I_{(1,0),j}\Big] = \mathbb{E}\Big[I_{(1),j}I_{(1,1,1),j}\Big] = O(\Delta t^2), \\
& \mathbb{E}I_{(1),j}^2 I_{(1,1),j}^2 = O(\Delta t^3).
\end{aligned}$$

Additional properties of stochastic integrals will be discussed during our analysis.

For a stochastic Itô integral $I_{\alpha_q,j}$, if index $\alpha_q$ contains only zeros, then this integral is deterministic. Then, if components of $\alpha_q$ are not all 0, the first moment of $I_{\alpha_q,j}$ is zero with probability 1 (see Lemma 5.7.1 of [14]). In the same subsection, authors provides a way to calculate the expectation of two stochastic Itô integrals (see Lemma 5.7.2 of [14]) which can be called second moment estimate.

We can see that stochastic integrals $I_{(1,0),j}$, $I_{(0,1),j}$, and $I_{(1,1,1),j}$ represent highest-order terms in the truncation (15). In particular, $\mathbb{E}[I_{(1,0),j}] = \mathbb{E}[I_{(0,1),j}] = \mathbb{E}[I_{(1,1,1),j}] = 0$ and $\mathbb{E}[I^2_{(1,0),j}] = \mathbb{E}[I^2_{(0,1),j}] = \mathbb{E}[I^2_{(1,1,1),j}] = O(\Delta t^3)$. Thus, these terms are of order $\Delta t^{3/2}$. Other stochastic integrals with three or more integral are of even higher order in $\Delta t$. Therefore, truncated higher integrals will not contribute to lower-order terms in the analysis of bias and MSE for the Drift and Diffusion estimators. We also keep term $I_{(0,0),j} = \Delta t^2/2$ which is formally of a higher order. The reasons are that (i) this term is deterministic does not present any difficulties in the analysis, and (ii) combined with the term $I_{(0),j} = \Delta t$, it can produce terms of order $\Delta t^3$, i.e. $I_{(0),j} I_{(0,0),j} = \Delta t^3/2$ which is the same order as $\mathbb{E}[I^2_{\alpha_q,j}]$ for $\alpha_q = (1,0), (0,1), (1,1,1)$.

### 2.6 Bias of the Drift term estimator

Recall our discrete estimator for the Drift term (10)

$$\hat{A}(x_k) = \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}(X_{t_j+\Delta t} - X_{t_j}).$$

Taking expectation of both sides for the expression above we obtain

$$\begin{aligned}
\mathbb{E}[\hat{A}(x_k)] &\approx \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\sum_{q=0}^{6}\mathbb{E}\left[A_q(X_{t_j})I_{\alpha_q,j}\right] \\
&= \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\sum_{q=0}^{6}\mathbb{E}_k\left[A_q(X_{t_j})\right]\mathbb{E}\left[I_{\alpha_q,j}\right] \\
&= \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\left(\mathbb{E}_k\left[A_0(X_{t_j})\right]\Delta t + \mathbb{E}_k\left[A_2(X_{t_j})\right]\frac{\Delta t^2}{2}\right) \\
&= \frac{1}{M}\sum_{j\in M_k}\mathbb{E}_k\left[A(X_{t_j})\right] + \frac{\Delta t}{2}\frac{1}{M}\sum_{j\in M_k}\mathbb{E}_k\left[A_2(X_{t_j})\right] \\
&= A(x_k) + C\,(\Delta x)^2 + O\left((\Delta x)^4\right) + O\,(\Delta t) \\
&= A(x_k) + O\left(\Delta x^2\right) + O\,(\Delta t) \\
&\longrightarrow A(x_k) \text{ as } \Delta x,\, \Delta t \to 0.
\end{aligned}$$

Here we used Lemma 2.5 to approximately compute conditional expectations where $\mathbb{E}_k[A(X_{t_j})]$ and $\mathbb{E}_k[A_2(X_{t_j})]$. Also, recall that according to the notation in (15) $A_0(x) \equiv A(x)$.

**Conclusion:** The Bias of the Drift term estimator behaves asymptotically as $\Delta t$, $\Delta x \to 0$ as

$$\mathbb{E}[\hat{A}(x_k)] - A(x_k) = O\left(\Delta x^2\right) + O\left(\Delta t\right).$$

## 2.7 Bias of the Diffusion term estimator

Recall the Diffusion estimator in (11)

$$\hat{B}^2(x_k) = \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\left(X(t_j+\Delta t) - X(t_j)\right)^2.$$

Taking expectation of both sides we obtain

$$\begin{aligned}
\mathbb{E}[\hat{B}^2(x_k)] \quad &= \quad \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\mathbb{E}_k\left(\sum_{q=0}^{6} A_q(X_{t_j}) I_{\alpha_q,j}\right)^2 \\
&= \quad \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\sum_{q,l=0}^{6}\mathbb{E}_k\left[A_q(X_{t_j})A_l(X_{t_j})\right]\mathbb{E}\left[I_{\alpha_q,j}I_{\alpha_l,j}\right] \\
&= \quad \frac{1}{M}\sum_{j\in M_k}\mathbb{E}_k\left[B^2(X_{t_j})\right] + \frac{1}{M}\sum_{j\in M_k}\mathbb{E}_k\left[A^2(X_{t_j})\right]\Delta t + \frac{C}{M}\sum_{j\in M_k}\left(\mathbb{E}_k\left[A_5^2(X_{t_j})\right]\right. \\
&\quad \left. +\mathbb{E}_k\left[B(X_{t_j})A_3(X_{t_j})\right] + \mathbb{E}_k\left[B(X_{t_j})A_4(X_{t_j})\right] + \mathbb{E}_k\left[A_1(X_{t_j})A_6(X_{t_j})\right]\right)\Delta t + O(\Delta t^2) \\
&= \quad \frac{1}{M}\sum_{j\in M_k}\mathbb{E}_k\left[B^2(X_{t_j})\right] + \tilde{C}\Delta t + O(\Delta t^2) \\
&= \quad B^2(x_k) + C\left(\Delta x\right)^2 + O\left(\left(\Delta x\right)^4\right) + O\left(\Delta t\right) \\
&= \quad B^2(x_k) + O\left(\Delta x^2\right) + O\left(\Delta t\right) \\
&\longrightarrow \quad B^2(x_k) \;\text{ as }\; \Delta x,\, \Delta t \to 0,
\end{aligned}$$

where we used Lemma 2.5 and properties of stochastic integrals in section 2.5. Also note that $A_1(x) \equiv B^2(x)$.

**Conclusion:** The Bias of the Diffusion term estimator behaves asymptotically as $\Delta t$, $\Delta x \to 0$

$$\mathbb{E}[\hat{B}^2(x_k)] - B^2(x_k) = O\left(\Delta x^2\right) + O\left(\Delta t\right).$$

## 2.8 Drawback of a slightly different estimator

We can define our estimators for the Drift and Diffusion coefficients slightly differently. Here we concentrate on the Drift estimator. In particular, when we introduce analog of conditional expectation for discrete summations, we can subtract the middle of the bin instead of the time-instance $X_{t_j}$, i.e., we can define the estimator for the Drift term as (compare with definition of $\hat{A}(x_k)$ in (10))

$$\tilde{A}(x_k) = \frac{1}{\Delta t}\frac{1}{M}\sum_{j \in M_k}\left(X_{t_j+\Delta t} - x_k\right) \tag{16}$$

where set $M_k$ is defined identically as for estimator (10). We will demonstrate that the definition above leads to an inferior estimator compared to $\hat{A}(x_k)$ in (10).

Taking expectation on both sides of (16) and using Lemma 2.5, we obtain the following expression for the bias

$$\begin{aligned}
\mathbb{E}[\tilde{A}(x_k)] &= \frac{1}{M\Delta t}\sum_{j \in M_k}\mathbb{E}_k\left[X_{t_j} - x_k\right] + \frac{1}{M\Delta t}\sum_{j \in M_k}\sum_{q=0}^{6}\mathbb{E}_k\left[A_q(X_{t_j})\right]\mathbb{E}\left[I_{\alpha_q,j}\right] \\
&= C\frac{(\Delta x)^2}{\Delta t} + \frac{1}{M}\sum_{j \in M_k}\mathbb{E}_k\left[A(X_{t_j})\right] + O(\Delta t) + O\left((\Delta x)^4/\Delta t\right) \\
&= C\frac{(\Delta x)^2}{\Delta t} + A(x_k) + O(\Delta t) + O\left(\Delta x^2\right) + h.o.t.,
\end{aligned}$$

where higher-order terms involve fractions $\Delta x^{2n}/\Delta t$ with $n \geq 2$.

Analyzing the bias for Drift term estimator $\tilde{A}(x_k)$ computed above, we find that even if we let $\Delta x, \Delta t \to 0$, this does not guarantee that this estimator is asymptotically unbiased because the ratio $\Delta x^2/\Delta t$ contributes to the bias of the Drift term estimator (16). Therefore, in the regime when $\Delta x^2/\Delta t \to c > 0$ Drift term estimator $\tilde{A}(x_k)$ will be biased for all $\Delta x$ and $\Delta t$. A similar conclusion can be reached for the modified Diffusion estimator. Therefore, $\tilde{A}(x_k)$ in (16) is inferior to the estimator $\hat{A}(x_k)$ introduced in (10).

### 2.9 MSE of Drift term estimator

It is well-known that consistency in the mean discussed in the sections 2.6 and 2.7 for the Drift and Diffusion estimators, respectively, does not imply convergence of these estimators to the true values of the Drift and diffusion. Moreover, we have three computational parameters - (i) the number of sampled points for computing discrete analog of expectation, $M$, (ii) sampling time-step, $\Delta t$, and (iii) size of each bin or space-discretization step, $\Delta x$. Thus, it is feasible that some combinations of these parameters (e.g., ratios and/or products) will determine the asymptotic behavior of estimators as $M \to \infty$ and $\Delta t$, $\Delta x \to 0$.

Therefore, a standard approach in the literature is to analyze the Mean-Square Error (MSE). The MSE is a widely used metric to analyze performance of estimators, since it takes both, the mean and variance of the estimator into account.

The MSE is defined as

$$MSE\{\hat{A}(x_k)\} = \mathbb{E}\left[\left(\hat{A}(x_k) - A(x_k)\right)^2\right] \tag{17}$$

and the expectation is taken with respect to the truncated density and distribution of the Brownian motion. If under some conditions $MSE\{\hat{A}(x_k)\} \to 0$, then the random variable $\hat{A}(x_k)$ converges to the constant $A(x_k)$ in $L^2$.

Recall the Drift estimator in (10)

$$\hat{A}(x_k) \quad = \quad \frac{1}{\Delta t}\frac{1}{M}\sum_{j \in M_k}(X_{t_j+\Delta t} - X_{t_j}).$$

Thus, we can substitute the truncated Itô-Taylor expansion (15) and the MSE becomes

$$\begin{aligned}
MSE\{\hat{A}(x_k)\} &= \mathbb{E}\left[\left(\hat{A}(x_k) - A(x_k)\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{M\Delta t}\sum_{j\in M_k}\left([A(X_{t_j}) - A(x_k)]\Delta t + \sum_{q=1}^{6} A_q(X_{t_j}) I_{\alpha_q,j}\right)\right)^2 \middle| X_{t_j} \in Bin_k\right] \\
&= \mathbb{E}\left[\frac{1}{M^2\Delta t^2}\sum_{i,j\in M_k}\left([A(X_{t_j}) - A(x_k)]\Delta t + \sum_{q=1}^{6} A_q(X_{t_j}) I_{\alpha_q,j}\right) \times \right. \\
&\qquad \left.\left([A(X_{t_i}) - A(x_k)]\Delta t + \sum_{l=1}^{6} A_l(X_{t_i}) I_{\alpha_l,i}\right) \middle| X_{t_i}, X_{t_j} \in Bin_k\right]. \qquad (18)
\end{aligned}$$

After unfolding formula (18), we have 49 terms in total. However, many terms are similar and, thus, we consider several groups of terms.

**Type 1:** Consider the cross-product of the first two terms in Formula (18) -

$$\begin{aligned}
&\frac{1}{(M\Delta t)^2}\sum_{i,j\in M_k}\left|\mathbb{E}\left[(A(X_{t_i}) - A(x_k))\,(A(X_{t_j}) - A(x_k))\Delta t^2 \middle| X_{t_i}, X_{t_j} \in Bin_k\right]\right| \\
&\le \frac{1}{M^2}\sum_{i,j\in M_k}\mathbb{E}\left[K_A^2\left|X_{t_i} - x_k\right|\left|X_{t_j} - x_k\right| \middle| X_{t_i}, X_{t_j} \in Bin_k\right] \\
&\le C\Delta x^2,
\end{aligned}$$

where we used that $A(x)$ is Lipschitz and $|X_{t_j} - x_k|$, $|X_{t_i} - x_k| \le \Delta x/2$ since both $X_{t_j}$, $X_{t_i} \in Bin_k$ and $x_k$ is the center of the bin.

**Type 2:** Consider cross-terms of the form

$$\frac{1}{(M\Delta t)^2}\sum_{i,j\in M_k}\mathbb{E}\Big[(A(X_{t_i})-A(x_k))\Delta t\, A_q(X_{t_j})I_{\alpha_q,j}+ \\ (A(X_{t_j})-A(x_k))\Delta t\, A_l(X_{t_i})I_{\alpha_l,i}\Big|X_{t_i},X_{t_j}\in Bin_k\Big] \\ =\frac{2}{M^2\Delta t}\sum_{i,j\in M_k}\mathbb{E}\left[(A(X_{t_i})-A(x_k))\,A_q(X_{t_j})I_{\alpha_q,j}\middle|X_{t_i},X_{t_j}\in Bin_k\right],\quad \text{for}\quad q=1,\ldots,6$$

where we used symmetry between $t_i$ and $t_j$ and $\alpha_l$ and $\alpha_q$. We have a total of 12 terms. The problem arises because $X_{t_i}$ and $I_{\alpha_l,j}$ are *not* independent if $t_i > t_j$. Thus, we cannot compute expectation above easily and proceed to simplify the expression above using the Lipschitz property of the Drift coefficient $A(x)$, Therefore, we use the Lipschitz property of $A(x)$ and obtain

$$\frac{2}{M^2\Delta t}\left|\mathbb{E}\left[\sum_q\sum_{i,j\in M_k}(A(X_{t_i})-A(x_k))A_q(X_{t_j})I_{\alpha_q,j}\middle|X_{t_i},X_{t_j}\in Bin_k\right]\right| \\ \le\frac{2}{M^2\Delta t}\sum_q\sum_{i,j\in M_k}\mathbb{E}\left[\left|\big(A(X_{t_i})-A(x_k)\big)\,A_l(X_{t_j})I_{\alpha_l,j}\right|\middle|X_{t_i},X_{t_j}\in Bin_k\right] \\ \le\frac{K_A\Delta x}{M^2\Delta t}\sum_q\sum_{i,j\in M_k}\mathbb{E}\left[\left|A_q(X_{t_j})I_{\alpha_q,j}\right|\middle|X_{t_j}\in Bin_k\right] \\ \le\frac{K_A\Delta x}{M^2\Delta t}\sum_q\sum_{i,j\in M_k}\left(\mathbb{E}_kA_q^2(X_{t_j})\right)^{1/2}\left(\mathbb{E}I^2_{\alpha_q,j}\right)^{1/2} \\ \le\frac{C\Delta x}{\sqrt{\Delta t}}\left(1+\sqrt{\Delta t}+O(\Delta t^{3/2})\right),$$

where we used the Hölder inequality in section 2.1 and lowest-order terms are due to $\mathbb{E}I^2_{(1),j}=\Delta t$ and $\mathbb{E}I^2_{(1,1),j}=\Delta t^2/2$. Other stochastic integrals contribute to higher-order terms. Here we use a notation for the conditional expectation $\mathbb{E}_kf(x)=\mathbb{E}[f(x)|x\in Bin_k]$. Since the truncated density has a finite support, we assume that all conditional expectations exist and are finite., e.g., $\mathbb{E}_kA_q^2(X_{t_j})<\infty$.

**Type 3:** Consider terms with stochastic integrals for either $q=2$ or $l=2$, in other words, we have the following 11 terms

$$\alpha_q = (0,0) \quad \text{and} \quad \alpha_l = (1),(1,1),(1,0),(0,1),(0,0),(1,1,1),$$
$$\alpha_l = (0,0) \quad \text{and} \quad \alpha_q = (1),(1,1),(1,0),(0,1),(1,1,1).$$

Due to symmetry, we only need to consider $q = 2$. Recall that $I_{(0,0),j} = \Delta t^2/2$. Then the formula (18) becomes

$$\begin{aligned} &\frac{1}{M^2}\left|\sum_{i,j\in M_k}\mathbb{E}\left[A_2(X_{t_i})A_q(X_{t_j})I_{\alpha_q,j}\middle|X_{t_i},X_{t_j}\in Bin_k\right]\right| \\ &\leq \frac{1}{M^2}\sum_{i,j\in M_k}\left(\mathbb{E}\left[A_2^2(X_{t_i})\,A_q^2(X_{t_j})\,\middle|X_{t_i},X_{t_j}\in Bin_k\right]\right)^{1/2}\|I_{\alpha_q,j}\|_2, \end{aligned}$$

where used the Hölder inequality in section 2.1. Similar to Type 2 terms, we assume that all expectations with respect to the joint truncated density exist and are finite (i.e. $\mathbb{E}_k\left[A_2^2(X_{t_i})\,A_q^2(X_{t_j})\right] < \infty$). The exact form of this joint density is hard to analyze, but it has a finite support and, thus, this assumption is quite reasonable.

As a final step, we only need to analyze lowest-order terms resulting from stochastic integrals. Therefore,

$$\begin{aligned} &\frac{1}{2M^2}\left|\sum_l\sum_{i,j\in M_k}\mathbb{E}\left[A_2(X_{t_i})\,A_q(X_{t_j})I_{\alpha_q,j}\,\middle|X_{t_i},X_{t_j}\in Bin_k\right]\right| \\ &\qquad \leq C\sqrt{\Delta t}\left(1+\sqrt{\Delta t}+O(\Delta t^{3/2})\right), \end{aligned}$$

where lowest-order terms are due to $\|I_{(1),j}\|_2 = \sqrt{\Delta t}$ and $\|I_{(1,1),j}\|_2 = \Delta t/\sqrt{2}$ and $C$ is some generic constant representing upper bound for all expectations of the form $\mathbb{E}_k\left[A_2^2(X_{t_i})\,A_q^2(X_{t_j})\right]$.

**Type 4:** Consider all possible combinations of stochastic integrals with the following indexes

$$\alpha_q, \alpha_l = (1),(1,1),(0,1),(1,0),(1,1,1). \tag{19}$$

From the property of stochastic integrals, we have $\mathbb{E}\left[A_q(X_{t_i})I_{\alpha_q,i}A_l(X_{t_j})I_{\alpha_l,j}\right] = 0$ when $i \neq j$. Assume that $t_j > t_i$. Then random variables $X_{t_i}$ and $X_{t_j}$ and increments of the Brownian motion on the interval $[t_j, t_j + \Delta t]$ are independent and, thus, $X_{t_i}$ and $X_{t_j}$ and stochastic integral $I_{\alpha_l,j}$ are independent random variables. Therefore,

$$\mathbb{E}\left[A_q(X_{t_i})I_{\alpha_q,i}A_l(X_{t_j})I_{\alpha_l,j}\right] = \mathbb{E}\left[A_q(X_{t_i})I_{\alpha_q,i}A_l(X_{t_j})\right]\mathbb{E}\left[I_{\alpha_l,j}\right] = 0 \quad \text{for} \quad t_j > t_i$$

from the properties of stochastic integral $I_{\alpha_l,j}$ for $l = 1, 3, 4, 5, 6$. A similar argument holds for $t_i > t_j$.

Therefore, for these combinations of stochastic integrals we only need to consider case $i = j$ and the terms with (19) in (18) become

$$\begin{aligned}
&\frac{1}{(M\Delta t)^2}\left|\sum_{i,j\in M_k}\mathbb{E}\left[A_q(X_{t_i})I_{\alpha_q,i}\,A_l(X_{t_j})I_{\alpha_l,j}\,\middle|X_{t_i}, X_{t_j} \in Bin_k\right]\right| \\
&= \frac{1}{(M\Delta t)^2}\left|\sum_{i\in M_k}\mathbb{E}\left[A_q(X_{t_i})\,A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}\,\middle|X_{t_i} \in Bin_k\right]\right|, \\
&\leq \frac{1}{(M\Delta t)^2}\sum_{i\in M_k}\left(\mathbb{E}_k\left[A_q^2(X_{t_i})\,A_l^2(X_{t_i})\right]\right)^{1/2}\|I_{\alpha_q,i}I_{\alpha_l,i}\|_2.
\end{aligned}$$

Either when $I_{\alpha_q,i} = I_{\alpha_l,i}$ or $I_{\alpha_q,i} \neq I_{\alpha_l,i}$, we need to calculate fourth moments of stochastic integrals. In particular, the lowest-order term is due to $\|I^2_{(1),i}\|_2 = \sqrt{3}\Delta t$. All other moments are of higher order, some of them are given by 5.2 and 5.7 of [14], Lemma 2.7 and (3)

$$\begin{aligned}
\|I_{(1),i}I_{(1,1),i}\|_2 &= O(\Delta t^{3/2}), \\
\|I^2_{(1,1),i}\|_2 &= \frac{\sqrt{15}}{2}(\Delta t)^2, \\
\|I^2_{(0,1),i}\|_2 &= \|I^2_{(1,0),i}\|_2 = \|I^2_{(1,1,1),i}\|_2 = O(\Delta t^3)
\end{aligned}$$

Therefore,

$$\frac{1}{(M\Delta t)^2}\left|\sum_{q,l}\sum_{i=1}^{M}\mathbb{E}\left[A_q(X_{t_i})\,A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}\left|X_{t_i}\in Bin_k\right.\right]\right|$$
$$\leq \frac{C}{M\Delta t}\left(1+\sqrt{\Delta t}+O(\Delta t)\right)$$

where summation over $q, l$ is taken over (19) and $C$ is a suitable constant.

### 2.9.1 Main results

Combining all terms we obtain

$$\boxed{MSE\{\hat{A}(x_k)\}\leq C\left(\sqrt{\Delta t}+\frac{1}{M\Delta t}+\Delta x^2+\frac{\Delta x}{\sqrt{\Delta t}}\right)+h.o.t.} \tag{20}$$

which describes the leading-order behavior of the Mean-Square Error for the Drift estimator. Here "*h.o.t.*" denotes higher-order terms.

Therefore, for the estimator to be asymptotically consistent, the following conditions need to be fulfilled

$$M\Delta t\to\infty,\quad \Delta t,\,\Delta x\to 0,\quad \Delta x/\sqrt{\Delta t}\to 0.$$

The first condition arises quite often in the analysis of various estimators for SDEs. The condition $M\Delta t \to \infty$ implies that the total observational time, $T$, should become infinite for the estimator to be consistent. The conditions $\Delta t$, $\Delta x \to 0$ are also not surprising, since these conditions were observed in the analysis of bias for the Drift estimator. This also could have been predicted from the definition of the Drift term (5) since this definition involves limit $\Delta t \to 0$ and it is reasonable that the size of the bin should tend to zero in order to mimic the conditional expectation in (5).

However, the last condition $\Delta x/\sqrt{\Delta t} \to 0$ is new and, thus, we discuss it in more detail. This condition implies that the bin size, $\Delta x$, has to tend to zero faster than $\sqrt{\Delta t}$. However, it also implies that $\Delta x$ might tend to zero slower than the observation time-step, $\Delta t$. To see this consider the scaling

$$\Delta x = \Delta t^a, \quad \text{then} \quad \frac{\Delta x}{\sqrt{\Delta t}} = \Delta t^{a-1/2} \to 0 \quad \text{for} \quad a > \frac{1}{2}.$$

However, since $\Delta t \to 0$, then if $\Delta x$ is chosen with $1/2 < a < 1$ then $\Delta t^a = \Delta x \gg \Delta t$. This has important practical consequences. In particular, this implies that the bin size can be chosen to be (much) larger compared to the observational time-step $\Delta t$. This, in turn, implies that it is much easier to obtain a trajectory with $M$ points in each bin because for larger bin size it is much easier to generate observational trajectories where $M$ points fall into a particular bin (interval).

To balance the error terms on the right-hand side of (20) we have to enforce $\sqrt{\Delta t} \sim \Delta x/\sqrt{\Delta t}$ which implies

$$\boxed{\Delta x \sim \Delta t.} \tag{21}$$

We also would like to point out that the scaling above effectively makes the error due to the $\Delta x^2$ term in (20) much smaller since $\Delta x^2 \ll \sqrt{\Delta t}$ for the scaling in (21). Thus, the error due to the larger bin size becomes negligible under the scaling (21).

Finally, we also would like to point out that the scaling in (21) also makes errors due to $O(\Delta x^2)$ negligible in the expression for the bias of $\hat{A}(x_k)$ considered in subsection 2.6. Recall, that $\mathbb{E}\hat{A}(x_k) = A(x_k) + O(\Delta x^2) + O(\Delta t)$.

To balance the first two error terms in the MSE of $\hat{A}(x_k)$ in (20) we can choose

$$\boxed{M \sim \Delta t^{-3/2}.} \tag{22}$$

The scalings (21) and (22) represent an optimal sampling regime for computing the Drift estimator $\hat{A}(x_k)$. We also expect that errors due to changes in the bin size should be negligible in the MSE of $\hat{A}(x_k)$.

To improve the computational efficiency of the Drift estimator one can choose the scaling

$$\Delta x \sim \Delta t^{0.5+\varepsilon}, \qquad 0 < \varepsilon \ll 1. \tag{23}$$

we would like to point out that the scaling above is not optimal, but the Drift term estimator is still consistent under this scaling. The scaling in (23) increases the bin size and potentially reduces the number of bins. Therefore, if the sampled trajectory is fixed, the scaling in (23) increased the number of points which falls into each bin and potentially reduces the error due to $(M\Delta t)^{-1}$ term.

### 2.10 MSE of Diffusion term estimator

In this section, we focus our attention on the performance of the Diffusion term estimator. In particular, we analyze the MSE

$$MSE\{\hat{B}^2(x_k)\} = \mathbb{E}\left[\left(\hat{B}^2(x_k) - B^2(x_k)\right)^2\right]. \tag{24}$$

After we substitute the truncated Itô-Taylor expansion (15) into the Diffusion term estimator (11), this estimator becomes

$$\hat{B}^2(x_k) \quad = \quad \frac{1}{\Delta t}\frac{1}{M}\sum_{j\in M_k}\left(\sum_{q=0}^{6} A_q(X_{t_j})I_{\alpha_q,j}\right)^2$$

and the MSE becomes

$$
\begin{aligned}
&MSE\{\hat{B}^2(x_k)\} \approx \mathbb{E}\left[\left(\frac{1}{M\Delta t}\sum_{j\in M_k}\left(\sum_{q=0}^{6} A_q(X_{t_j})I_{\alpha_q,j}\right)^2 - B^2(x_k)\right)^2\right] \\
&= \mathbb{E}\left[\left(\frac{1}{M\Delta t}\sum_{j\in M_k}\sum_{q,l=0}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j} - B^2(x_k)\right)^2\right] \\
&= \underbrace{\frac{1}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k}\left(B^2(X_{t_i})I^2_{(1),i} - B^2(x_k)\Delta t\right)\left(B^2(X_{t_j})I^2_{(1),j} - B^2(x_k)\Delta t\right)\right]}_{Type1} + \qquad (25) \\
&\quad \underbrace{\frac{1}{(M\Delta t)^2}\mathbb{E}\left[\left(\sum_{j\in M_k}\sum_{\substack{q,l=0\\ q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right)^2\right]}_{Type2} + \qquad (26) \\
&\quad \underbrace{\frac{2}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k}\left(B^2(X_{t_i})I^2_{(1),i} - B^2(x_k)\Delta t\right)\times \sum_{\substack{q,l=0\\ q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right]}_{Type3}. \qquad (27)
\end{aligned}
$$

Similar to the MSE for the Drift estimator, we consider several different types of terms in the expression above, but first, we state a simple lemma

**Lemma 2.8.**

$$\mathbb{E}_k\left[B^2(X_{t_i})B^2(X_{t_j})\left(I^2_{(1),i} - \Delta t\right)\left(I^2_{(1),j} - \Delta t\right)\right] = 0, \quad \textit{for } i \neq j.$$

*Proof.* Because of symmetry, we only consider $j > i$. In particular, consider $j = i + k$ with $k > 0$. Since time internals $[t_i, t_i + \Delta t]$ and $[t_{i+k}, t_{i+k} + \Delta t]$ do not overlap (because $t_i + \Delta t \leq t_{i+1}$)

$$\mathbb{E}\left[B^2(X_{t_i})B^2(X_{t_{i+k}})\left(I^2_{(1),i} - \Delta t\right)\left(I^2_{(1),i+k} - \Delta t\right)\right]$$

$$= \mathbb{E}\left[B^2(X_{t_i})B^2(X_{t_{i+k}})\left(I^2_{(1),i} - \Delta t\right)\right]\mathbb{E}\left[\left(I^2_{(1),i+k} - \Delta t\right)\right] = 0$$

because $\mathbb{E}\left[\left(I^2_{(1),i+k} - \Delta t\right)\right] = 0.$ □

Next we separate all the terms in the MSE of the Diffusion estimator into 3 different types.

**Type 1:** First, we consider the first term in (25) and by adding and subtracting $B^2(X_{t_i})\Delta t$ and $B^2(X_{t_j})\Delta t$ in the first and second bracket, respectively, we obtain

$$\begin{aligned}
&\frac{1}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k}\left(B^2(X_{t_i})I^2_{(1),i} - B^2(x_k)\Delta t\right)\left(B^2(X_{t_j})I^2_{(1),j} - B^2(x_k)\Delta t\right)\right] \\
&= \frac{1}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})B^2(X_{t_j})\left(I^2_{(1),i} - \Delta t\right)\left(I^2_{(1),j} - \Delta t\right)\right] \\
&\quad + \frac{2}{M^2\Delta t}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})\left(I^2_{(1),i} - \Delta t\right)\left(B^2(X_{t_j}) - B^2(x_k)\right)\right] \\
&\quad + \frac{1}{M^2}\mathbb{E}\left[\sum_{i,j\in M_k}\left(B^2(X_{t_i}) - B^2(x_k)\right)\left(B^2(X_{t_j}) - B^2(x_k)\right)\right] \\
&\le \frac{1}{(M\Delta t)^2}\sum_{i\in M_k}\mathbb{E}_k\left[B^4(X_{t_i})\right]\mathbb{E}\left[\left(I^2_{(1),i} - \Delta t\right)^2\right] + C\frac{K_B\Delta x}{\Delta t}\mathbb{E}\left[\left|I^2_{(1),i} - \Delta t\right|\right] + \frac{(K_B\Delta x)^2}{4} \\
&= C\left(\frac{1}{M} + \Delta x + \Delta x^2\right).
\end{aligned}$$

where $C$ is some constant and $K_B$ is a Lipschitz constant for $B^2(x)$.

**Type 2:** Consider the terms arising from (26). There are many terms come from squaring the sum in (26), but all of them have the following form

$$\frac{1}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} A_q(X_{t_i})A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}A_r(X_{t_j})A_m(X_{t_j})I_{\alpha_r,j}I_{\alpha_m,j}\right], \tag{28}$$

where $q, l, r, m = 0, \cdots, 6$ with restriction $q \times l \neq 1$ and $r \times m \neq 1$ since the case $q \times l = 1$ and $r \times m = 1$ corresponds to terms of type 1 and type 3 considered separately. This means that we cannot have $\alpha_q = \alpha_l = (1)$ or $\alpha_r = \alpha_m = (1)$. Here indexes $q, l$ correspond to time $t_i$ and indexes

$r, m$ correspond to time $t_j$. There are total of $M^2$ terms in the summation in (28) . There are many terms of this type, and we will distinguish several sub-types.

**Type 2(i):**

Consider type 2 terms with the following 3 restrictions -

– 1) at least one of the integrals in each pair $I_{\alpha_q,i}I_{\alpha_l,i}$ and $I_{\alpha_r,j}I_{\alpha_m,j}$ is stochastic,

– 2) $q \neq l$ and $r \neq m$,

– 3) the number of 1's in both $(\alpha_q, \alpha_l)$ and $(\alpha_r, \alpha_m)$ pairs are odd.

Without loss of generality we can consider

$$\mathbb{E}\left[A_q(X_{t_i})A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}A_r(X_{t_j})A_m(X_{t_j})I_{\alpha_r,j}I_{\alpha_m,j}\right] \quad \text{with} \quad j > i.$$

Using $j > i$, we can write

$$\begin{aligned}
&\mathbb{E}\left[A_q(X_{t_i})A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}A_r(X_{t_j})A_m(X_{t_j})I_{\alpha_r,j}I_{\alpha_m,j}\right] \\
&= \mathbb{E}\left[A_q(X_{t_i})A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}A_r(X_{t_j})A_m(X_{t_j})\right]\mathbb{E}\left[I_{\alpha_r,j}I_{\alpha_m,j}\right]
\end{aligned}$$

Since time intervals $[t_i, t_i + \Delta t]$ and $[t_j, t_j + \Delta t]$ do not overlap. And using the fact that the number of 1's in the pair $(\alpha_r, \alpha_m)$ is odd,

$$\mathbb{E}\left[I_{\alpha_r,j}I_{\alpha_m,j}\right] = 0.$$

where we use Lemma 5.7.2 in [14]. A similar argument holds for $i > j$. Therefore, since the number of 1's in both pairs $I_{\alpha_q,i}I_{\alpha_l,i}$ and $I_{\alpha_r,j}I_{\alpha_m,j}$ is odd, we can reduce the sum in (28) to the case $i = j$, i.e.,

$$\begin{aligned}
&\frac{1}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} A_q(X_{t_i})A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}A_r(X_{t_j})A_m(X_{t_j})I_{\alpha_r,j}I_{\alpha_m,j}\right] \\
&= \frac{1}{(M\Delta t)^2}\mathbb{E}\left[\sum_{j\in M_k} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}A_r(X_{t_j})A_m(X_{t_j})I_{\alpha_r,j}I_{\alpha_m,j}\right].
\end{aligned}$$

This reduces the number of terms in the summation from $M^2$ (when $i, j \in M_k$) to $M$ (when $j \in M_k$).

There are many combinations of indexes $\alpha_q, \alpha_l, \alpha_r, \alpha_m$ which satisfy requirements for Type 2(i) terms. Intuitively, it is clear that since $I_{(1),i} \sim \sqrt{\Delta t}$ and $I_{(1),j} \sim \sqrt{\Delta t}$ are lowest-order stochastic integrals in the expansion (15), lowest-order terms for Type 2(i) will appear when $q = r = 1$ (or when $l = m = 1$ by symmetry). In this case $\alpha_q = \alpha_r = (1)$. Let us remind about the restriction we're considering, if $q = r = 1$, then $l \neq 1$ and $m \neq 1$ due to restriction 2) for Type 2(i) above. This significantly reduces the number of terms. In addition, it is also clear that in order to capture the leading-order terms of Type 2(i) integrals $I_{\alpha_l,i}$ and $I_{\alpha_m,j}$ should be of the lowest possible order. There are two integrals of order $\Delta t$, namely $I_{(0),i} = \Delta t$ and $||I_{(1,1),i}||_2 \sim \Delta t$. Therefore, indexes $\alpha_l$ and $\alpha_m$ should correspond to those two integrals. Thus, here we list some lower order terms of Type 2(i).

(a) $q = r = 1$ and $l = m = 5$ or we can switch $q$, $l$ and $r$, $m$ because of symmetry. In this case $\alpha_q = \alpha_r = (1)$ and $\alpha_l = \alpha_m = (1, 1)$.

$$
\begin{aligned}
&\frac{1}{(M\Delta t)^2} \sum_{i,j \in M_k} \mathbb{E}\left[A_1(X_{t_i})A_5(X_{t_i})I_{(1),i}I_{(1,1),i}A_1(X_{t_j})A_5(X_{t_j})I_{(1),j}I_{(1,1),j}\right] \\
&= \frac{1}{(M\Delta t)^2} \sum_{j \in M_k} \mathbb{E}\left[A_1^2(X_{t_j})A_5^2(X_{t_j})I_{(1),j}^2 I_{(1,1),j}^2\right] \\
&= \frac{1}{M\Delta t^2} \mathbb{E}_k\left[B_1^2(x)B_2^2(x)\right] \|I_{(1),j}^2 I_{(1,1),j}^2\|_2^2 \leq \frac{C\Delta t}{M}.
\end{aligned}
$$

(b) $q = r = 1$ and $l = m = 0$ or we can switch $q$, $l$ and $r$, $m$. In this case $\alpha_q = \alpha_r = (1)$ and $\alpha_l = \alpha_m = (0)$ and we would like to remind that $I_{(0),i} = I_{(0),j} = \Delta t$. Therefore, (28) reduces to

$$
\begin{aligned}
&\frac{1}{M^2} \sum_{i,j \in M_k} \mathbb{E}\left[A_1(X_{t_i})A_0(X_{t_i})I_{(1),i}A_1(X_{t_j})A_0(X_{t_j})I_{(1),j}\right] \\
&= \frac{1}{M^2} \sum_{j \in M_k} \mathbb{E}\left[A_1^2(X_{t_j})A_0^2(X_{t_j})I_{(1),j}^2\right] \\
&= \frac{1}{M^2} \sum_{j \in M_k} \mathbb{E}_k\left[A_1^2(X_{t_j})A_0^2(X_{t_j})\right] \mathbb{E}\left[I_{(1),j}^2\right] \leq \frac{C\Delta t}{M}.
\end{aligned}
$$

(c) $q=r=1$, $l=0$ and $m=5$ (index $q$ corresponds to time $t_i$) or we can switch $q$, $l$ and $r$, $m$. In this case $I_{\alpha_l,i}=I_{(0),i}=\Delta t$, $I_{\alpha_m,j}=I_{(1,1),j}$ and, therefore, (28) becomes

$$
\begin{aligned}
&\frac{1}{(M\Delta t)^2}\sum_{i,j\in M_k}\mathbb{E}\left[A_1(X_{t_i})A_0(X_{t_i})I_{(1),i}\Delta t A_1(X_{t_j})A_5(X_{t_j})I_{(1),j}I_{(1,1),j}\right]\\
&=\frac{1}{M^2\Delta t}\sum_{j\in M_k}\mathbb{E}\left[A_1^2(X_{t_j})A_0(X_{t_j})A_5(X_{t_j})I_{(1),j}^2 I_{(1,1),j}\right]\\
&=\frac{1}{M^2\Delta t}\sum_{j\in M_k}\mathbb{E}_k\left[A_1^2(X_{t_j})A_0(X_{t_j})A_5(X_{t_j})\right]\mathbb{E}\left[I_{(1),j}^2 I_{(1,1),j}\right]\leq\frac{C\Delta t}{M}.
\end{aligned}
$$

where we utilize Lemma 2.7 and (3) to expand the expectation term $\mathbb{E}\left[I_{(1),j}^2 I_{(1,1),j}\right]$, then we use Lemma 5.7.2 and Lemma 5.7.5 in [14] to obtain the order of $\Delta t$. Other terms result in higher-order terms. Therefore, Type 2(i) terms are equivalent to $O(\Delta t/M)$.

**Type 2(ii):**

Consider type 2 terms with the following 3 restrictions -

– 1) at least one of the integrals in each pair $I_{\alpha_q,i}I_{\alpha_l,i}$ and $I_{\alpha_r,j}I_{\alpha_m,j}$ is stochastic,

– 2) $q\neq l$ and $r\neq m$,

– 3) the number of 1's in $(\alpha_q,\alpha_l)$ or $(\alpha_r,\alpha_m)$ pairs is even.

The main difference between type 2(ii) and type 2(i) is that for type 2(ii) the number of 1's in either $(\alpha_q,\alpha_l)$ or $(\alpha_r,\alpha_m)$ is even. This condition is complimentary to the condition 3) of type 2(i) terms. Here we cannot reduce the summation over $i,j\in M_k$, and, therefore, there will be $M^2$ terms in the summation $\sum\limits_{i,j\in M_k}$. Here we have to provide different types of estimates compared with Type 2(i) terms.

In particular, we consider

$$
\begin{aligned}
&\frac{1}{(M\Delta t)^2}\left|\mathbb{E}\left[\sum_{i,j\in M_k}A_q(X_{t_i})A_l(X_{t_i})I_{\alpha_q,i}I_{\alpha_l,i}A_r(X_{t_j})A_m(X_{t_j})I_{\alpha_r,j}I_{\alpha_m,j}\right]\right|\\
&\leq\frac{1}{\Delta t^2}\|A_q(X_{t_i})A_l(X_{t_i})A_r(X_{t_j})A_m(X_{t_j})\|_2\,\|I_{\alpha_q,i}I_{\alpha_l,i}I_{\alpha_r,j}I_{\alpha_m,j}\|_2\\
&=\frac{C}{\Delta t^2}\|I_{\alpha_q,i}I_{\alpha_l,i}I_{\alpha_r,j}I_{\alpha_m,j}\|_2. \qquad (29)
\end{aligned}
$$

We would like to point out that when $i \neq j$ the $L^2$ norm above reduces to

$$\|I_{\alpha_q,i}I_{\alpha_l,i}I_{\alpha_r,j}I_{\alpha_m,j}\|_2 = \|I_{\alpha_q,i}I_{\alpha_l,i}\|_2\,\|I_{\alpha_r,j}I_{\alpha_m,j}\|_2 \quad \text{for } i \neq j.$$

The constant $C$ is always finite because we're computing the expectation and the norm with respect to the joint conditional distribution of $X_{t_i}$ and $X_{t_j}$, i.e.,

$$\begin{aligned}&\|A_q(X_{t_i})A_l(X_{t_i})A_r(X_{t_j})A_m(X_{t_j})\|_2\\ &= \left(\mathbb{E}\left[\Big(A_q(X_{t_i})A_l(X_{t_i})A_r(X_{t_j})A_m(X_{t_j})\Big)^2 | X_{t_i}, X_{t_j} \in Bin_k\right]\right)^{1/2}.\end{aligned}$$

Therefore, we have to compute the lowest-order terms of the form

$$\|I_{\alpha_q,i}I_{\alpha_l,i}I_{\alpha_r,j}I_{\alpha_m,j}\|_2 \tag{30}$$

where we can use Lemma 5.7.5 in [14] as $i = j$ for estimates of higher moments belong to a multiple Ito integral and Lemma 5.7.2 in [14] as $i \neq j$ for second moment estimates with indexes restricted to Type 2(ii). We would like to note that without any restrictions, the lowest order terms would be $\|I^2_{(1),i}I^2_{(1),j}\|_2 = \Delta t^2$. However, with restrictions for Type 2(ii) outlined above neither $I^2_{(1),i}$ nor $I^2_{(1),j}$ are allowed.

Without the loss of generality we consider the case when number of 1's in the pair $(\alpha_r, \alpha_m)$ is even. We just list the lowest order terms:

(a) When $(\alpha_q, \alpha_l) = ((1),(0))$ and $(\alpha_r, \alpha_m) = ((0),(1,1))$, we have

$$\Delta t^{-2}\|I_{\alpha_q,i}I_{\alpha_l,i}I_{\alpha_r,j}I_{\alpha_m,j}\|_2 = \Delta t^{-2}\Delta t^2\|I_{(1),i}I_{(1,1),j}\|_2 \leq C\Delta t^{\frac{3}{2}}$$

for both, $i = j$ and $i \neq j$. In fact, the norm above can be computed exactly in both cases using $I_{\alpha_{(1)},i} = \Delta W_{i+1}$ and Lemma 2.7 and (3).

(b) When $(\alpha_q, \alpha_l) = ((1), (0))$ and $(\alpha_r, \alpha_m) = ((1), (0, 1))$, we have

$$\Delta t^{-2} \|I_{\alpha_q,i} I_{\alpha_l,i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = \Delta t^{-2} \Delta t \|I_{(1),i} I_{(1,1),j} I_{(0,1),j}\|_2 \leq C \Delta t^{\frac{3}{2}}$$

for both $i = j$ and $i \neq j$. Where we use Lemma 2.7 and (3), ignore the higher order terms when $i = j$. Considering $i \neq j$, the stochastic integrals with different subscripts are independent, we can seperate them into the multiplication of two $L^2$ norm, according to Lemma 5.7.2 in [14].

(c) When $(\alpha_q, \alpha_l) = ((1), (0))$ and $(\alpha_r, \alpha_m) = ((1), (1, 0))$, we have

$$\Delta t^{-2} \|I_{\alpha_q,i} I_{\alpha_l,i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = \Delta t^{-2} \Delta t \|I_{(1),i} I_{(1,1),j} I_{(1,0),j}\|_2 \leq C \Delta t^{\frac{3}{2}}$$

for both $i = j$ and $i \neq j$. Similar reasons of calculation as (b), If $i = j$, we utilize Lemma 2.7 and (3). If $i \neq j$, we use independence and Lemma 5.7.2 in [14].

(d) When $(\alpha_q, \alpha_l) = ((1), (1, 1))$ and $(\alpha_r, \alpha_m) = ((0), (1, 1))$, we have

$$\Delta t^{-2} \|I_{\alpha_q,i} I_{\alpha_l,i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = \Delta t^{-1} \|I_{(1),i} I_{(1,1),i} I_{(1,1),j}\|_2 \leq C \Delta t^{\frac{3}{2}}$$

for both $i = j$ and $i \neq j$. When $i = j$, we use direct computation by Lemma 2.7. If $i \neq j$, we had independence, Lemma 2.7, and (3).

(e) When $(\alpha_q, \alpha_l) = ((1), (1, 1))$ and $(\alpha_r, \alpha_m) = ((1), (0, 1))$, we have

$$\Delta t^{-2} \|I_{\alpha_q,i} I_{\alpha_l,i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = \Delta t^{-2} \|I_{(1),i} I_{(1,1),i} I_{(1),j} I_{(0,1),j}\|_2 \leq C \Delta t^{\frac{3}{2}}$$

for both $i = j$ and $i \neq j$. Considering $i = j$, we use (3), Lemma 2.7 and Lemma 5.7.2 in [14]. If $i \neq j$, we have independence, Lemma 2.7 and (3).

(f) When $(\alpha_q, \alpha_l) = ((1), (1, 1))$ and $(\alpha_r, \alpha_m) = ((1), (1, 0))$, we have

$$\Delta t^{-2} \|I_{\alpha_q,i} I_{\alpha_l,i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = \Delta t^{-2} \|I_{(1),i} I_{(1,1),i} I_{(1),j} I_{(1,0),j}\|_2 \leq C \Delta t^{\frac{3}{2}}$$

for both $i = j$ and $i \neq j$. Where we use exactly the same reasons as bounding (e), because they have different $\alpha_m$.

(g) When $(\alpha_q, \alpha_l) = ((1), (1, 1))$ and $(\alpha_r, \alpha_m) = ((1), (1, 1, 1))$, we have

$$\Delta t^{-2} \| I_{\alpha_q,i} I_{\alpha_l,i} I_{\alpha_r,j} I_{\alpha_m,j} \|_2 = \Delta t^{-2} \| I_{(1),i} I_{(1,1),i} I_{(1),j} I_{(1,1,1),j} \|_2 \leq C \Delta t^{\frac{3}{2}}$$

for both $i = j$ and $i \neq j$. From 5.2.21 of [14], we have expression $I_{(1,1,1),j} = \frac{1}{3!}(I^3_{(1),j} - 3\Delta t I_{(1),j})$. We use this with (3) and Lemma 2.7 for calculation in case $i = j$. For case $i \neq j$, we have independence, Lemma 2.7, and Lemma 5.7.2 in [14].

Therefore, Type 2(ii) terms are equivalent to $O(\Delta t^{3/2})$.

**Type 2(iii):**

Here we consider the case when both integrals in the same pair $I_{\alpha_q,i} I_{\alpha_l,i}$ or $I_{\alpha_r,j} I_{\alpha_m,j}$ are deterministic. Without the loss of generality we consider both integrals $I_{\alpha_q,i} I_{\alpha_l,i}$ to be deterministic. There are only two determined integrals considered in the truncated Ito-Taylor expansion (15), namely $I_{(0),i} = \Delta t$ and $I_{(0,0),i} = \Delta t^2/2$. Clearly, the lowest-order terms arise from $(\alpha_q, \alpha_l) = ((0), (0))$ (i.e., $I_{\alpha_q,i} I_{\alpha_l,i} = I^2_{(0),i} = \Delta t^2$).

Here we use the same reduction as for Type 2(ii) terms in (29). In particular, we write

$$\begin{aligned}
&\frac{1}{(M\Delta t)^2} \left| \mathbb{E} \left[ \sum_{i,j \in M_k} A_q(X_{t_i}) A_l(X_{t_i}) I_{\alpha_q,i} I_{\alpha_l,i} A_r(X_{t_j}) A_m(X_{t_j}) I_{\alpha_r,j} I_{\alpha_m,j} \right] \right| \\
&\leq C \| I_{\alpha_r,j} I_{\alpha_m,j} \|_2.
\end{aligned}$$

Here we list some lowest-order terms.

(a) When $(\alpha_q, \alpha_l) = ((0), (0))$ and $(\alpha_r, \alpha_m)$ is $((1), (0))$ or $((0), (1))$, we have

$$\| I_{\alpha_r,j} I_{\alpha_m,j} \|_2 = \Delta t \| I_{(1),j} \|_2 = \Delta t^{3/2}.$$

(b) When $(\alpha_q, \alpha_l) = ((0),(0))$ and $(\alpha_r, \alpha_m)$ is $((1),(1,1))$ or $((1,1),(1))$, we have

$$\|I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = \|I_{(1),j} I_{(1,1),j}\|_2 \leq C\Delta t^{3/2}.$$

All other combinations of stochastic integrals yield terms of higher order.

Therefore, Type 2(iii) terms are equivalent to $O(\Delta t^{3/2})$.

**Type 2(iv):**

Consider terms with $q = l$ or $r = m$. Without loss of generality we consider the case $q = l$. We would like to remind that Type 2 terms are computed under the restriction $q \times l \neq 1$, which means that we cannot have $\alpha_q = \alpha_l = (1)$. The next stochastic integral which yield the lowest-order terms is $q = l = 5$ or $I^2_{(1,1),i} \sim \mathbb{E}\left[\Delta W^4_{i+1}\right] \sim \Delta t^2$.

Here we use the same reduction as for Type 2(ii) terms in (29). In particular, we write

$$\begin{aligned}
&\frac{1}{(M\Delta t)^2}\left|\mathbb{E}\left[\sum_{i,j\in M_k} A_q(X_{t_i}) A_l(X_{t_i}) I_{\alpha_q,i} I_{\alpha_l,i} A_r(X_{t_j}) A_m(X_{t_j}) I_{\alpha_r,j} I_{\alpha_m,j}\right]\right| \\
&\leq \frac{C}{\Delta t^2}\|I^2_{(1,1),i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2.
\end{aligned}$$

The lowest-order terms arise from combination of indexes $(\alpha_r, \alpha_m = ((1),(0))$ and $(\alpha_r, \alpha_m = ((1),(1,1))$.

(a) When $(\alpha_r, \alpha_m = ((1),(0))$ we have

$$C\Delta t^{-2}\|I^2_{(1,1),i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = C\Delta t^{-2}\Delta t\|I^2_{(1,1),i} I_{(1),j}\|_2 \leq C\Delta t^{3/2}.$$

for both $i = j$ and $i \neq j$. Where we use Lemma 2.7 and (3).

(b) When $(\alpha_r, \alpha_m = ((1),(1,1))$ we have

$$C\Delta t^{-2}\|I^2_{(1,1),i} I_{\alpha_r,j} I_{\alpha_m,j}\|_2 = C\Delta t^{-2}\Delta t\|I^2_{(1,1),i} I_{(1),j} I_{(1,1),j}\|_2 \leq C\Delta t^{3/2}.$$

for both $i = j$ and $i \neq j$. Where we use Lemma 2.7 and (3). All other combinations of integrals yield terms of higher order.

Therefore, Type 2(iv) terms are equivalent to $O(\Delta t^{3/2})$.

**Type 3:** Finally, we consider Type 3 terms and using Lemma 2.7 we obtain

$$\frac{2}{(M\Delta t)^2}\mathbb{E}\left[\left(\sum_{i,j\in M_k}\left(B^2(X_{t_i})I^2_{(1),i} - B^2(x_k)\Delta t\right)\sum_{\substack{q,l=0\\q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right)\right]$$

$$= \frac{2}{(M\Delta t)^2}\mathbb{E}\left[\left(\sum_{i,j\in M_k}B^2(X_{t_i})\left(I^2_{(1),i} - \Delta t\right)\sum_{\substack{q,l=0\\q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right)\right] +$$

$$\frac{2}{(M\Delta t)^2}\mathbb{E}\left[\left(\sum_{i,j\in M_k}\Delta t\left(B^2(X_{t_i}) - B^2(x_k)\right)\sum_{\substack{q,l=0\\q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right)\right]$$

$$= \underbrace{\frac{4}{(M\Delta t)^2}\mathbb{E}\left[\left(\sum_{i,j\in M_k}B^2(X_{t_i})I_{(1,1),i}\sum_{\substack{q,l=0\\q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right)\right]}_{Type3(i)} +$$

$$\underbrace{\frac{2}{M^2\Delta t}\mathbb{E}\left[\left(\sum_{i,j\in M_k}\left(B^2(X_{t_i}) - B^2(x_k)\right)\sum_{\substack{q,l=0\\q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right)\right]}_{Type3(ii)}.$$

**Type 3(i):** For the first term in Type 3(i), the lowest-order terms arise from $q = 1$, $l = 0$ and $q = 1$, $l = 5$ which corresponds to $(\alpha_q, \alpha_l) = ((1),(0))$ and $(\alpha_q, \alpha_l) = ((1),(1,1))$, respectively.

(a) Consider $q = 1$, $l = 0$ first. Then using the same argument as in Lemma 2.8 we can show that

$$\begin{aligned}
&\frac{4}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})I_{(1,1),i}A_1(X_{t_j})A_0(X_{t_j})I_{(1),j}\Delta t\right] \\
&= \quad \frac{4}{M^2\Delta t}\mathbb{E}\left[\sum_{i\in M_k} B^2(X_{t_i})I_{(1,1),i}A_1(X_{t_i})A_0(X_{t_i})I_{(1),i}\right] \\
&= \quad \frac{4}{M^2\Delta t}\sum_{i\in M_k}\mathbb{E}_k\left[B^2(X_{t_i})A_1(X_{t_i})A_0(X_{t_i})\right]\mathbb{E}\left[I_{(1,1),i}I_{(1),i}\right] = 0.
\end{aligned}$$

(b) Next, consider $q = 1$, $l = 5$. Then using the same argument as in Lemma 2.8 we obtain

$$\begin{aligned}
&\frac{4}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})I_{(1,1),i}A_1(X_{t_j})A_5(X_{t_j})I_{(1),j}I_{(1,1),j}\right] \\
&= \quad \frac{4}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i\in M_k} B^2(X_{t_i})I_{(1,1),i}A_1(X_{t_i})A_5(X_{t_i})I_{(1),i}I_{(1,1),i}\right] \\
&= \quad \frac{4}{(M\Delta t)^2}\sum_{i\in M_k}\mathbb{E}_k\left[B^2(X_{t_i})A_1(X_{t_i})A_5(X_{t_i})\right]\mathbb{E}\left[I_{(1),i}I^2_{(1,1),i}\right] = 0.
\end{aligned}$$

(c) The next order terms appear due to combinations of indexes which correspond to $(\alpha_q, \alpha_l) = ((1), (0, 1))$, $(\alpha_q, \alpha_l) = ((1), (1, 0))$, and $(\alpha_q, \alpha_l) = ((1), (1, 1, 1))$. In these cases we cannot apply argument from Lemma 2.8 since $\mathbb{E}\left[I_{(1),i}I_{(1,0),i}\right] \neq 0$, $\mathbb{E}\left[I_{(1),i}I_{(0,1),i}\right] \neq 0$, and $\mathbb{E}\left[I_{(1),i}I_{(1,1,1),i}\right] \neq 0$. Therefore, we proceed as in (29) to obtain

$$\begin{aligned}
&\frac{4}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})I_{(1,1),i}A_1(X_{t_j})A_4(X_{t_j})I_{(1),j}I_{(1,0),j}\right] \\
&\leq \frac{4}{\Delta t^2}\|B^2(X_{t_i})A_1(X_{t_j})A_4(X_{t_j})\|_2\,\|I_{(1,1),i}I_{(1),j}I_{(1,0),j}\|_2 \leq C\Delta t.
\end{aligned}$$

A similar argument can be applied to $(\alpha_q, \alpha_l) = ((1), (0, 1))$ and $(\alpha_q, \alpha_l) = ((1), (1, 1, 1))$.

(d) We would like to point out that the combination of indexes $(\alpha_q, \alpha_l) = ((0), (1, 1))$ yield a higher-order term because in this case we can use the argument similar to Lemma 2.8, i.e.,

$$
\begin{aligned}
&\frac{4}{(M\Delta t)^2}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})I_{(1,1),i}A_0(X_{t_j})A_5(X_{t_j})\Delta t I_{(1,1),j}\right] \\
&= \frac{4}{M^2\Delta t}\mathbb{E}\left[\sum_{i,j\in M_k} B^2(X_{t_i})I_{(1,1),i}A_0(X_{t_j})A_5(X_{t_j})I_{(1,1),j}\right] \\
&= \frac{4}{M^2\Delta t}\mathbb{E}\left[\sum_{i\in M_k} B^2(X_{t_i})A_0(X_{t_i})A_5(X_{t_i})I^2_{(1,1),i}\right] \\
&\le \frac{4}{M^2\Delta t}\sum_{i\in M_k}\mathbb{E}_k\left[B^2(X_{t_i})A_0(X_{t_i})A_5(X_{t_i})\right]\mathbb{E}\left[I^2_{(1,1),i}\right] \le \frac{C\Delta t}{M}.
\end{aligned}
$$

**Type 3(ii):**

$$
\begin{aligned}
&\frac{2}{(M\Delta t)^2}\left|\mathbb{E}\left[\sum_{i,j\in M_k}\Delta t\left(B^2(X_{t_i})-B^2(x_k)\right)\sum_{\substack{q,l=0\\ q\times l\neq 1}}^{6} A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right]\right| \\
&\le \frac{2K_d\Delta x}{\Delta t}\sum_{\substack{q,l=0\\ q\times l\neq 1}}^{6}\mathbb{E}\left[\left|A_q(X_{t_j})A_l(X_{t_j})I_{\alpha_q,j}I_{\alpha_l,j}\right|\right] \\
&= \frac{2K_B\Delta x}{\Delta t}\sum_{\substack{q,l=0\\ q\times l\neq 1}}^{6}\mathbb{E}_k\left[\left|A_q(x)A_l(x)\right|\right]\|I_{\alpha_q,j}I_{\alpha_l,j}\|_1.
\end{aligned}
$$

When $q=1$, $l=0$ or $q=1$, $l=5$, we get the lowest order term. Consider $q=1$, $l=0$. Then $\|I_{(1),j}I_{(0),j}\|_1 = \Delta t\|I_{(1),j}\|_1 = O(\Delta t^{3/2})$ and

$$
\frac{2K_d\Delta x}{\Delta t}\mathbb{E}_k\left[\left|A_1(x)A_0(x)\right|\right]\|I_{(1),j}I_{(0),j}\|_1 \le C\Delta x\sqrt{\Delta t}.
$$

One can also show that $\|I_{(1),j}I_{(1,1),j}\|_1 \sim \mathbb{E}[|\Delta W^3_{j+1}|] = O(\Delta t^{3/2})$ which yields a similar bound for $q=1$, $l=5$.

### 2.10.1 Main results

$$\boxed{MSE\{\hat{B}^2(x_k)\} \leq C\left(\frac{1}{M} + \Delta x + \Delta t\right) + h.o.t.} \tag{31}$$

where $C$ is some constant.

# 3 Numerical Simulations

In this chapter we present numerical simulations to check our analytical results discussed in Chapter 2. In particular, we apply our estimators in (10) and (11) to simulated data and analyze the performance of these estimators in various parameter regimes. In Chapter 2, we derived asymptotic formulas for the MSEs of the Drift and Diffusion estimators. These asymptotic formulas have three parameters - (i) the sub-sampling (or observational) time-step $\Delta t$, (ii) the number $M$ of observed data-points in each bin, (iii) the space-discretization $\Delta x$. The expressions for MSEs of $\hat{A}(x)$ and $\hat{B}^2(x)$ contain various combinations of these three parameters and, therefore, different terms in the expressions for MSEs become dominant depending on different behavior of computational parameters $\Delta t$, $M$, and $\Delta x$. The goal of this chapter is to verify the asymptotic behavior of different errors terms and study numerically how errors from different terms dominate or balance each other.

In practice some error terms can be small and some error terms can be quite large. Since error terms are closely interwoven, we consider two important cases - (i) $M\Delta t \to \infty$ and (ii) $M\Delta t = Const$. The expression $T = M\Delta t$ represents the total time of the sampled trajectory required for estimation. Thus, the two regimes mentioned above represent two different sampling schemes - for $T \to \infty$ the total time-length of the trajectory required for estimation of the Drift and Diffusion coefficients diverges to infinity, and for $T = Const$ the length of the trajectory stays finite. In addition, our analytical estimates involve various "generic constants" which arise through various inequalities and bounds. It is impossible to determine the precise value of these constants and elucidate the most dominant error terms. Thus, we use the numerical results to study the importance of different terms in the MSEs of the Drift and Diffusion estimators. In addition, in practice, it is important to develop a estimation strategy which is the least computationally expensive. Therefore, it is important to assess the balance of various error terms and understand how changes in computational parameters reduce or increase the overall estimation errors.

We use two sets of simulations with two different stochastic processes in this Chapter. In particular, we perform the simulations of the Ornstein-Uhlenbeck (OU) process and the nonlinear (cubic) model with multiplicative noise.

## 3.1 Two particular examples for numerical simulations

In this section, we present two different stochastic processes – the Ornstein-Uhlenbeck process (OU) and the Nonlinear Drift and Multiplicative Noise (cubic) process. For the OU process, we utilize a hybrid numerical scheme where we use a second order Runge-Kutta scheme for the deterministic part and Euler discretization for the noise. For the cubic process, we use 1.5 order strong Itô-Taylor scheme for robust and accurate simulations [14].

### 3.1.1 Ornstein-Uhlenbeck process

In this section we present the Ornstein-Uhlenbeck process, describe numerical scheme, and list parameters in the numerical simulations. The Ornstein-Uhlenbeck process is given by the following stochastic differential equation

$$dX_t = -\gamma X_t dt + \sigma dW_t. \tag{32}$$

We use a second order Runge-Kutta discretization for the deterministic part and Euler scheme for the stochastic part

$$\begin{aligned} K_1 &= f(x), \qquad K_2 = f\left(x + \frac{\delta t}{2}, K_1\right), \qquad f(x) = -\gamma x, \\ X_{t+\delta t} &= X_t + K_2(X_t)\delta t + \sigma \delta W, \end{aligned}$$

where $\delta W \sim N(0, \delta t)$.

We choose the following parameters in our simulations

$$\gamma = 0.5, \qquad \sigma = 1$$

and time-step $\delta t = 0.0005$. We define intervals (bins) for the estimation and conditioning as in (9) where $\Delta x$ is the parameter representing the size of the bin and $x_k$ is the center of the bin. Results presented here will correspond to estimating the Drift $A(x)$ and Diffusion $B^2(x)$ on a finite interval $[-L, L]$ with $L = 1$. This is motivated by the stationary variance of the OU process for the above choice of $\gamma$ and $\sigma$

$$Stationary\ Var\{X_t\} = \frac{\sigma^2}{2\gamma} = 1.$$

We would like to point out that the Ornstein-Uhlenbeck process is a Gaussian process. Moreover, since the Drift term is linear like $A(x) = -\gamma x$ and the Diffusion term is a constant like $B(x) = \sigma$. Given the Drift and Diffusion terms, we have some coefficients for higher-order terms in the Itô-Taylor expansion (15) are zero (e.g. $A_3 \equiv A_5 \equiv A_6 \equiv 0$). Therefore, our generic bounds for the behavior of the bias and MSE can be re-derived taking explicitly into account absence of those terms. Therefore, the generic asymptotic behavior discussed in the previous chapter can be quite different compared to a particular case of the OU process. Nevertheless, we do not repeat the derivation of the bias and MSE for the Ornstein-Uhlenbeck process separately. Instead, we observe numerically that the behavior of the bias and MSEs for this process is rather generic and agrees with the overall asymptotic behavior of MSEs for a more general nonlinear process considered in the next section.

### 3.1.2 Nonlinear drift and Multiplicative Noise (Cubic) process

In this section we present the "cubic" process, numerical scheme and parameters. We consider the following stochastic process with cubic drift and linear Diffusion and refer to this process as the cubic process

$$dX_t = -\gamma X_t^3 dt + (\sigma_1 + \sigma_2 X_t) dW_t. \tag{33}$$

In our numerical simulations we utilize the order 1.5 strong Itô-Taylor discretization [14]

$$
\begin{aligned}
X_{t+\delta t} \;=\;& X_t - \gamma X_t^3 \delta t + (\sigma_1 + \sigma_2 X_t)\delta W + \frac{\sigma_2(\sigma_1 + \sigma_2 X_t)}{2}((\delta W)^2 - \delta t) - 3\gamma X_t^2(\sigma_1 + \sigma_2 X_t)\delta Z \\
& + \frac{1}{2}\left(3\gamma^2 X_t^5 - 3\gamma X_t(\sigma_1 + \sigma_2 X_t)^2\right)\delta t^2 - \gamma \sigma_2 X_t^3(\delta W \delta t - \delta Z) \\
& + \frac{\sigma_2^2(\sigma_1 + \sigma_2 X_t)\delta W}{2}\left(\frac{1}{3}(\delta W)^3 - \delta t\right),
\end{aligned}
$$

where $\delta W \sim N(0, \delta t)$, $\delta Z$ is an approximation for $I_{(1,0)} = \int_t^{t+\Delta t} \int_t^s dW(t')ds$ and the pair of random variables $(\delta W, \delta Z)$ can be determined from two independent random variables $U_1 \sim N(0,1)$ and $U_2 \sim N(0,1)$ by a linear transformation [14]. $(\delta W, \delta Z)$ can be computed explicitly and have the following properties

$$
\delta W = U_1 \sqrt{\delta t}, \qquad \delta Z = \frac{(\delta t)^{\frac{3}{2}}}{2}\left(U_1 + \frac{1}{\sqrt{3}}U_2\right),
$$

with

$$
\mathbb{E}[\delta Z] = 0, \quad \mathbb{E}[(\delta Z)^2] = \frac{1}{3}\delta t^3, \quad \mathbb{E}[\delta Z \delta W] = \frac{1}{2}\delta t^2.
$$

We choose the following parameters in our simulations

$$
\gamma = 1, \quad \sigma_1 = \sigma_2 = \frac{1}{\sqrt{2}}, \quad \delta t = 0.0005.
$$

For the above choice of parameters

$$
A(x) = -x^3, \qquad B^2(x) = 0.5x^2 + x + 0.5.
$$

We define the estimation bin as in (9) where $\Delta x$ is the parameter representing the size of the bin, and $x_k$ is the center of the bin. We estimate the Drift and Diffusion terms on the interval $[-L, L]$ with $L = 0.5$ since the variance of the cubic process is smaller than the variance of the OU process. For the cubic process $Stationary\ Var\{X_t\} \approx 0.25$.

We would like to point out that the cubic process in (33) represents a more generic situation for studying numerically the behavior of the Drift and Diffusion estimators since terms in the Itô-Taylor expansion are not zero.

## 3.2 Conditional moments

In this section, we analyze numerically the behavior of moments with respect to the truncated density. In particular, we analyze numerically the convergence of moments as $\Delta x \to 0$. The corresponding analytical expressions were derived in section 2.4.1, Lemma 2.5. In particular, we showed that the conditional moments obey

$$\mathbb{E}[X_t^p | \mathbb{1}(X_t \in Bin_k)] = x_k^p + O(\Delta x^2).$$

For the both, Ornstein-Uhlenbeck (32) and cubic (33) processes coefficients in the Itô-Taylor expansion become polynomial functions. Therefore, we need to verify the behavior of non-central conditional moments. The property above was used in the derivation of our generic bounds for the bias and MSE of the Drift and Diffusion estimators. Therefore, we would like to verify numerically that the bin size is selected adequately and we're working in the correct regime (i.e., the expression above is approximately correct).

### 3.2.1 Conditional moments of the Ornstein-Uhlenbeck process

We present the simulation results with changing $\Delta x$ to verify the above formula. We consider $M = 1000$ points which fall in each bin and the discrete estimation formula becomes

$$\mathbb{E}[X_t^p | \mathbb{1}(X_t \in Bin_k)] \approx \frac{1}{M} \sum_{i=1}^{M} x_i^p \mathbb{1}(x_i \in Bin_k) \equiv m^p(k), \tag{34}$$

where $p$ is the order of the corresponding moment. We select the following values for $\Delta x$

$$\Delta x = 0.01,\ 0.025,\ 0.05,\ 0.1$$

and the estimation interval $x_k \in [-L, L]$ with $L = 1$.

To analyze numerically the behavior of moments with respect to the truncated density we compute the Absolute error

$$\text{AbsErr}_p = \max_k \left| m^p(k) - x_k^p \right|, \tag{35}$$

where $m^p(k)$ is defined in (34) and $x_k$ is the center of the $k$th bin.

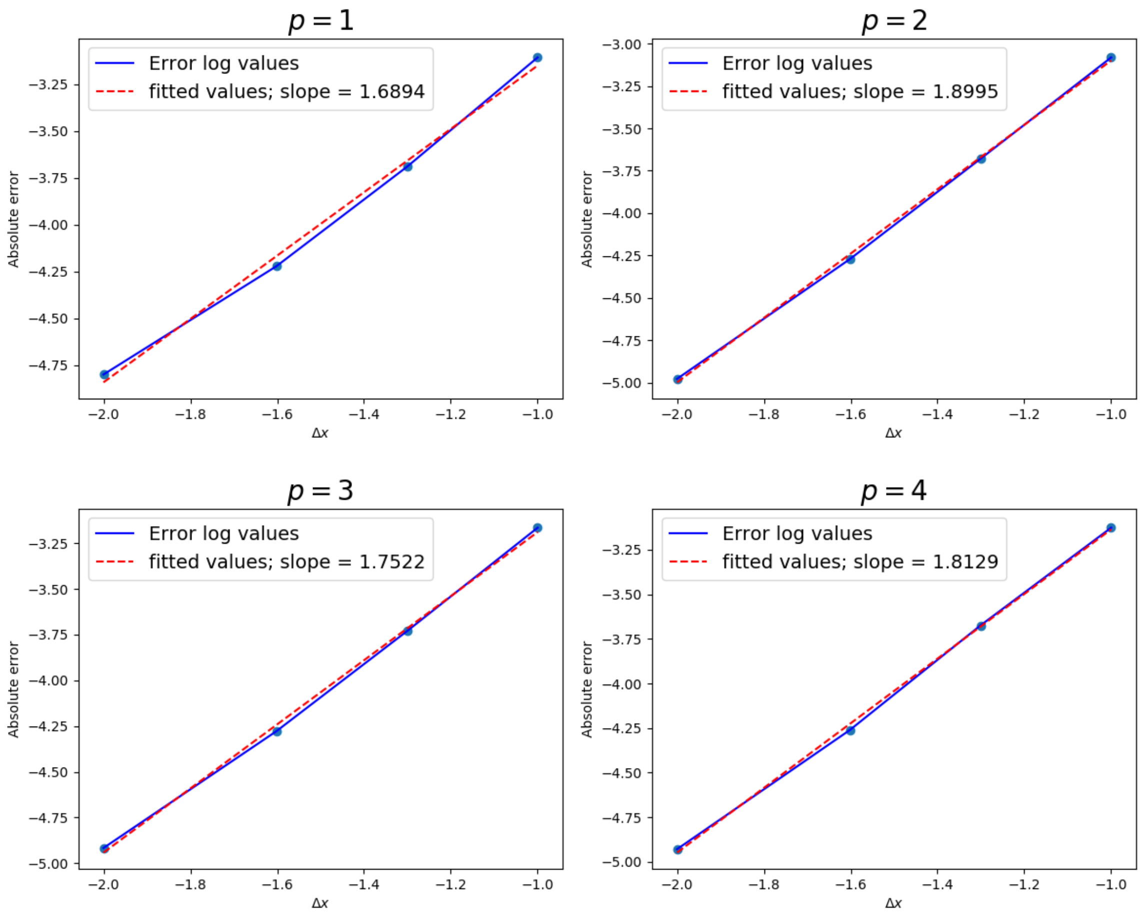



Figure 1: Log-log plot of the absolute error in the estimation of conditional moments $\mathbb{E}[X_t^p | \mathbb{1}(X_t \in Bin_k)]$ with $p = 1, \ldots, 4$ as defined in (35). Solid Blue line - numerically computed errors, Dashed Red line - linear fit.

Figure 1 presents the log-log plots of $\text{AbsErr}_p$ vs $\Delta x$ for $p = 1, \ldots, 4$. This plot shows that for moments up to order four, the absolute error decays approximately as $O(\Delta x^2)$. In particular, the slopes for linear fits are 1.6894, 1.8995, 1.7522, 1.8128.

### 3.2.2 Conditional moments of the cubic process

In this section we verify numerically the behavior of conditional moments for cubic model (33). We consider $M = 1000$ points in each bin and the discrete estimation formulas (34), where $p$ is the order of the corresponding moment. We select the following values for $\Delta x$

$$\Delta x = 0.005,\ 0.0125,\ 0.025,\ 0.05$$

and the estimation interval $x_k \in [-L, L]$ with $L = 0.5$.

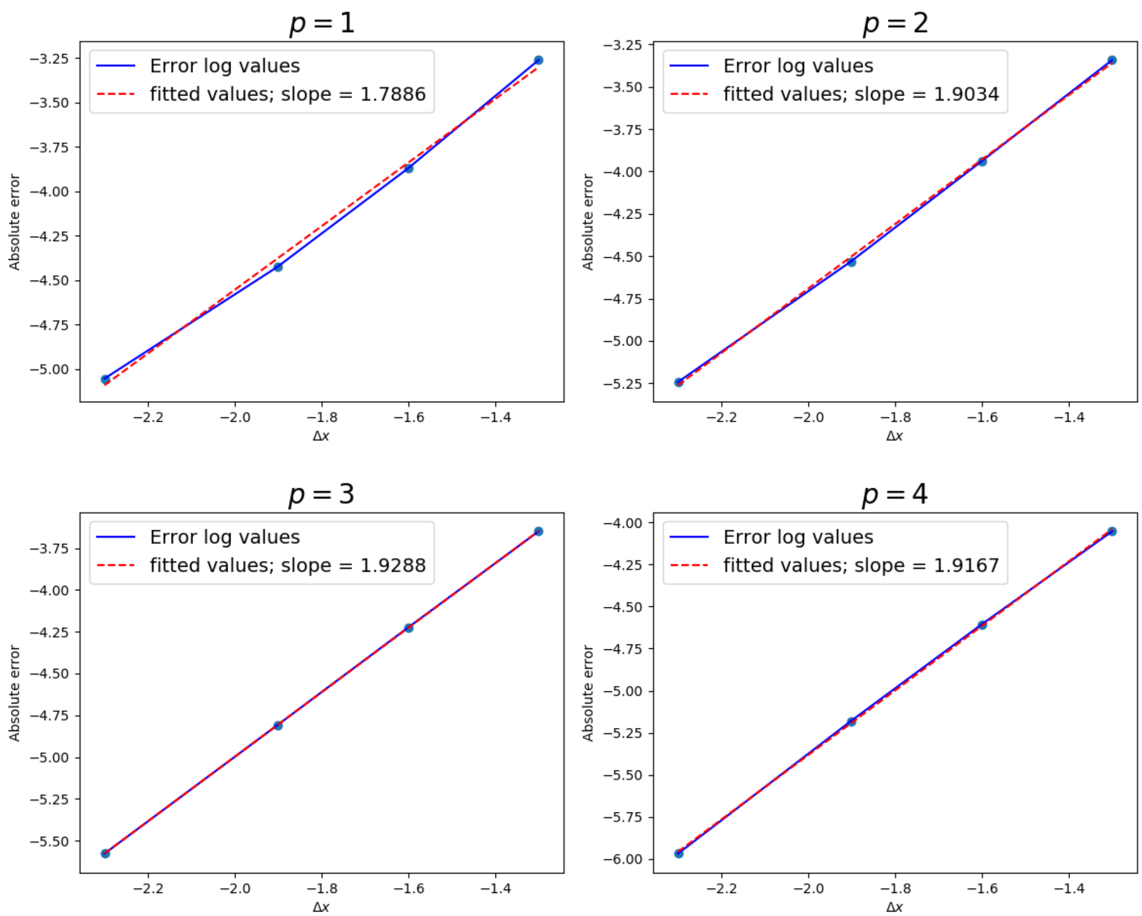



Figure 2: Log-log plot of the absolute error in the estimation of conditional moments $\mathbb{E}[X_t^p|\mathbb{1}(X_t \in Bin_k)]$ with $p = 1, \ldots, 4$ as defined in (35). Solid line - numerically computed errors, Dashed line - linear fit.

Similar to the previous section we plot the absolute error defined in (35). The results for the convergence of conditional moments are presented in Figure 2. Similar to the simulations for the OU process, Figure 2 shows that the absolute error decays approximately as $O(\Delta x^2)$. In particular, the slopes for linear fits are 1.7886, 1.9034, 1.9288, 1.9167.

### 3.2.3 Conclusions

In this section we verified numerically that for bin sizes in the range of

$$\Delta x \in [0.005, 0.1]$$

the conditional moments behave according to the analytical prediction derived in Lemma 2.5. Therefore, this range of bin sizes appears to be suitable to the estimation of the Drift and Diffusion coefficients and verification of our bounds for MSEs for the corresponding estimators. In practice, we would like to keep the bin size as large as possible, since this will allow to collect more points for estimation and, thus, reduce the computational complexity. The bin size $\Delta x \approx 0.1$ seems to be adequate for this purpose. We'll perform addition investigation how bin size affects the estimation in subsequent sections.

## 3.3 Absolute errors for the Drift and Diffusion estimators

In this section, we analyze numerically the absolute errors for the drift and Diffusion estimation. We analyze several regimes when $M \to \infty$ and $\Delta t \to 0$. The goal of this section is to provide the numerical evidence for the behavior of the absolute errors of the Drift and Diffusion estimators and obtain practical guidelines for selecting the computational parameters $M$, $\Delta t$, and $\Delta x$. Although we do not have the analytical expressions for the absolute errors, they are often used as a measure of accuracy for the estimators. Thus, we carry out a numerical investigation of the absolute errors as we vary the estimation parameters $\Delta t$, $M$, and $\Delta x$. We define absolute errors for Drift and

Diffusion terms as

$$AbsErr_{drift} = \frac{1}{MC}\sum_{j=1}^{MC}\left(\max_k \left|\hat{A_k}^{(j)} - A(k)\right|\right), \tag{36}$$

$$AbsErr_{diff} = \frac{1}{MC}\sum_{j=1}^{MC}\left(\max_k \left|\hat{B_k^2}^{(j)} - B^2(k)\right|\right), \tag{37}$$

and $MC$ is the number of Monte-Carlo Realizations, $\hat{A_k}^{(j)}$ and $\hat{B_k^2}^{(j)}$ are the estimates for the Drift and Diffusion value for the $j$th Monte-Carlo Realization, respectively. The parameter $k$ represents the $k$th bin.

### 3.3.1 Ornstein-Uhlenbeck process

For the OU process the Drift and Diffusion terms are given by

$$A(k) = -\gamma x_k, \qquad B^2(k) = \sigma^2,$$

with the parameters described in section 3.1.1. We consider the number of points in each bin

$$M = 50, 100, 200, 500, 1000 \tag{38}$$

and we also perform three sets of runs with

$$\Delta x = 2L/20,\ 2L/40,\ 2L/80, \quad x_k \in [-L, L], \quad L = 1. \tag{39}$$

The choice of $L$ is motivated by the stationary variance of the OU process $\sigma^2/(2\gamma) = 1$. Therefore, we effectively sample the points for estimation in the range $[-StdDev, StdDev]$ which ensures that there is enough points in each bin. We simulate $MC = 500$ realizations and consider two estimation regimes $M\Delta t \to \infty$ and $M\Delta t = Const$. For $M\Delta t = Const$ we use $\Delta t = 0.02, 0.01, 0.005, 0.002, 0.001$ for the corresponding value of $M$ in (38) and for $M\Delta t \to \infty$ we use

$\Delta t = 0.01$ for all values of $M$ in (38).

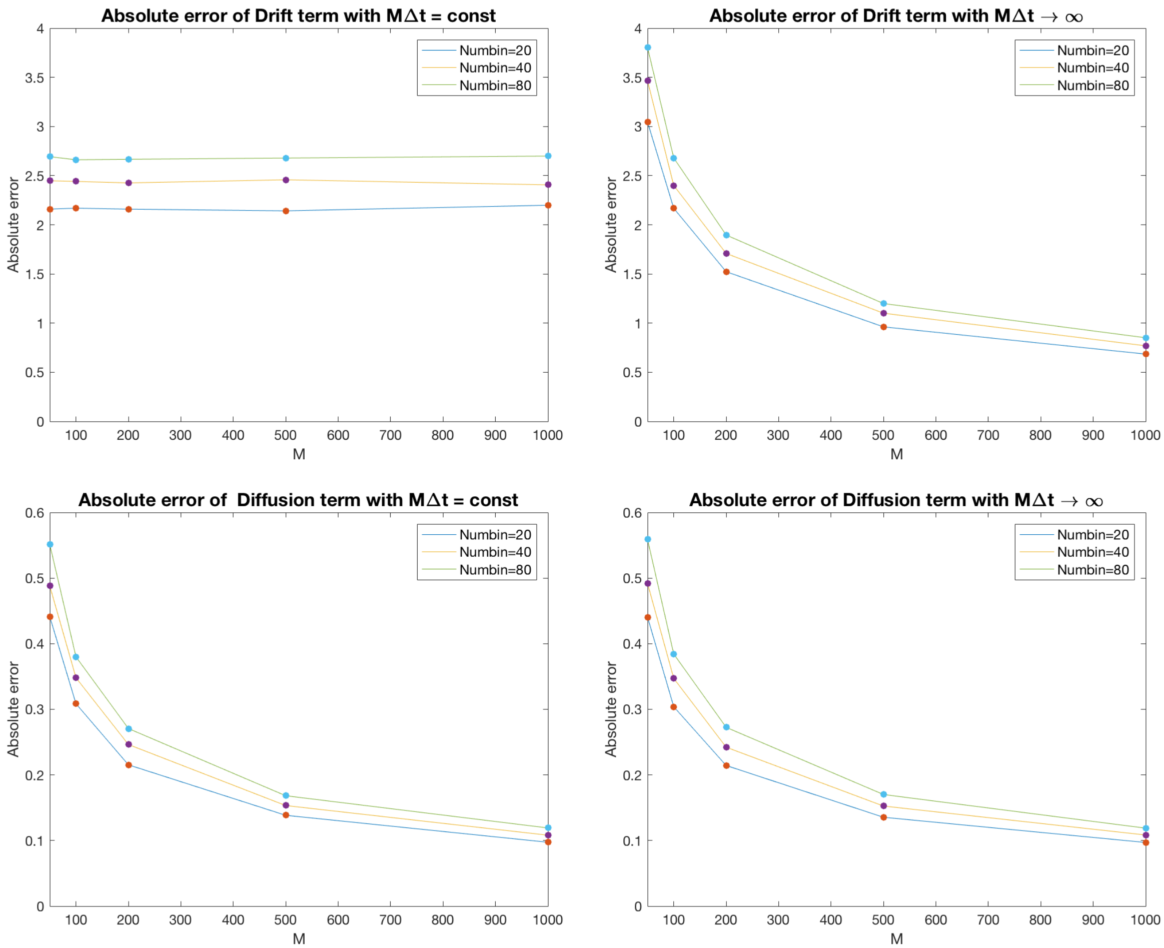



Figure 3: Absolute errors of the Drift (top) and Diffusion (bottom) estimators for the OU process with two different sampling regimes $M\Delta t = 500$ (left) and $M\Delta t \to \infty$ (right).

Figure 3 depicts the behaviors of absolute errors in both regimes for the Drift and Diffusion estimators. For the case $M\Delta t = Const$, the absolute errors for the Drift estimator are approximately constant as $M \to \infty$. When $M\Delta t \to \infty$, the top-right part of Figure 3 demonstrates that errors decrease sharply as $M\Delta t \to \infty$ (with fixed $\Delta t$). This reflects our analytic prediction for the MSE in (20). We can see that the behaviors of absolute errors for the Diffusion estimator is very similar in two regimes and is primarily driven by the $O(M^{-1})$ term which agrees with the analytical prediction for the MSE of the Diffusion estimator in (31).

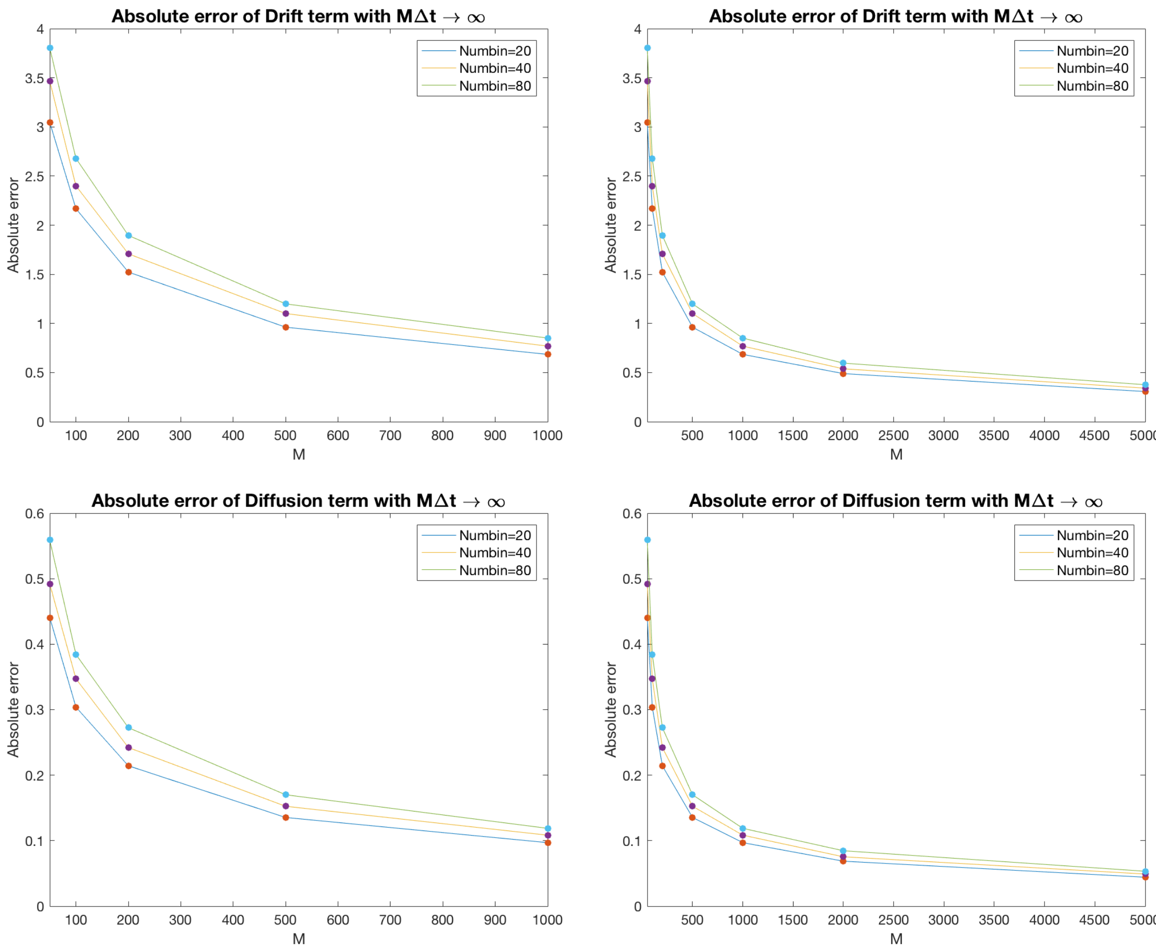



Figure 4: Absolute errors of Drift (top) and Diffusion (bottom) estimators for the OU process for $M\Delta t \to \infty$ for two different ranges of $M \in [50, 1000]$ (left) and $M \in [50, 5000]$ (right).

Larger number of observational points, $M$.

We can see in Figure 3 that for the range of $M$ considered in those simulations ($M \in [50, 1000]$) the absolute errors decay significantly in this range, and there is a significant decrease of the absolute errors from $M = 500$ to $M = 1000$ in the regime $M\Delta t \to \infty$ (right part of Figure 3). Thus, the terms $(M\Delta t)^{-1}$ and $M^{-1}$ for errors of the Drift and Diffusion estimators, respectively, seem to be significant for this range of $M$. Therefore, we extend the range of $M$ and consider

$$M = 50, 100, 200, 500, 1000, 2000, 5000 \qquad \text{fixed } \Delta t = 0.01 \tag{40}$$

to analyze numerically the range where terms $O(M^{-1})$ are significant. This corresponds to the case $M\Delta t \to \infty$. The numerical results are presented in Figure 4 and Table 1. We can see that the absolute errors are still quite large in the range $M \in [50, 5000]$, especially for the Drift estimator.

Table 1: Absolute errors of Drift and Diffusion estimators for the OU process for $M\Delta t \to \infty$ on $M \in [50, 5000]$ for three cases $NumBin = 20$, $NumBin = 40$, and $NumBin = 80$ where $\Delta x = 2L/NumBin$.

| Absolute error | | | | | | |
|---|---|---|---|---|---|---|
| | Drift estimator | | | Diffusion estimator | | |
| $M$ | $\Delta x = 0.1$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ | $\Delta x = 0.1$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ |
| 50 | 3.045 | 3.466 | 3.803 | 0.4405 | 0.4919 | 0.5588 |
| 100 | 2.168 | 2.400 | 2.679 | 0.3039 | 0.3471 | 0.3842 |
| 200 | 1.523 | 1.708 | 1.896 | 0.2143 | 0.2422 | 0.2725 |
| 500 | 0.9621 | 1.101 | 1.200 | 0.1355 | 0.1526 | 0.1701 |
| 1000 | 0.6850 | 0.7691 | 0.8510 | 0.09702 | 0.1084 | 0.1188 |
| 2000 | 0.4895 | 0.5382 | 0.5962 | 0.06872 | 0.07544 | 0.08468 |
| 5000 | 0.3071 | 0.3415 | 0.3764 | 0.04424 | 0.04910 | 0.05336 |

Smaller $\Delta x$ and $\Delta t$.

Figure 4 demonstrates that the absolute errors are quite large, especially for the Drift estimator. Even for large values of $M = 5000$, the absolute error is still approximately $AbsError\{Drift\} \approx 0.3, \ldots, 0.38$ (see Table 1). This is an indication that error terms due to other computational parameters (i.e. $\Delta x$ and $\Delta t$) become significant for the larger values of $M$. Therefore, we consider the estimation regime with much smaller $\Delta x$ and $\Delta t$ and larger $M$

$$M = 5000, 10000, 25000 \tag{41}$$

with the time steps $\Delta t = 0.002$ and $\Delta x = 2L/160$. The numerical results in this regime are presented in Figure 5 (cf., with Figure 4 for $\Delta t = 0.01$ and $\Delta x = 2L/80$). Please note that Figures are both depicted on the same vertical scale. **Drift Estimator.** First, we compare the absolute error for the Drift Estimator for $M = 5000$ and $(\Delta t, \Delta x) = (0.002, 2L/160)$ vs. $(\Delta t, \Delta x) = (0.01, 2L/80)$ (cf., the left part of Figure 5 and the top-right part of Figure 4). We observe that the absolute error *increases* for $(\Delta t, \Delta x) = (0.002, 2L/160)$. This suggests that there

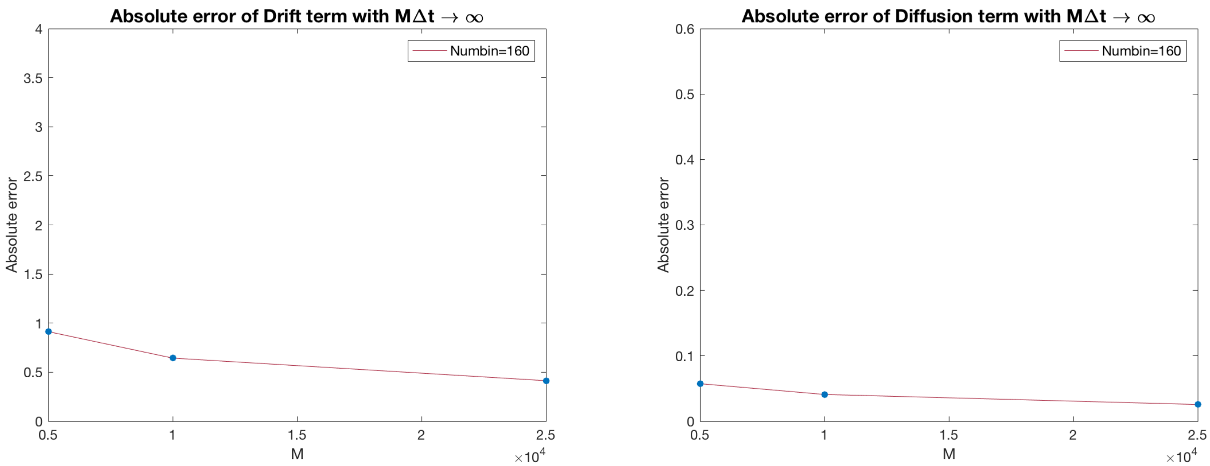



Figure 5: Absolute errors of the Drift and Diffusion term estimators for the OU process with smaller $\Delta t = 0.002$ and $\Delta x = 2L/160$ and $M$ in (41).

is a term with $\Delta t$ in the denominator. This is consistent with our analytical estimate for MSE in (20). It is rather difficult to estimate the precise form of this error term and, in particular, the power of $\Delta t$ in the denominator because there are also the other error terms which contribute to the overall error increase/decrease. However, this computational example suggests that it is not practical to select a very small $\Delta t$, since it may yield large errors in the estimation of the Drift term, even for very large $M$. **Diffusion Estimator.** Next, we compare the absolute error for the Diffusion Estimator for $M = 5000$ and $(\Delta t, \Delta x) = (0.002, 2L/160)$ vs. $(\Delta t, \Delta x) = (0.01, 2L/80)$ (cf., the right part of Figure 5 and the bottom-right part of Figure 4). We observe that smaller computational parameters $\Delta t$ and $\Delta x$ do not affect significantly the absolute error computed with $M = 5000$ points. Therefore, we can conclude that it is unlikely that there is a term with $\Delta t$ in the denominator, which is consistent with our analysis in (31).

### 3.3.2 Cubic process

For the cubic process (33) the Drift and Diffusion coefficients are given by

$$A(k) = -\gamma x_k^3, \qquad B^2(k) = (\sigma_1 + \sigma_2 x_k)^2.$$

We consider the parameter values described in section 3.1.2. We consider values of $M$ in (38) and we also perform three sets of runs with

$$\Delta x = 2L/20,\ 2L/40,\ 2L/80, \quad x_k \in [-L, L], \quad L = 0.5. \tag{42}$$

We consider $MC = 500$ and two estimation regimes $M\Delta t = Const$ and $M\Delta t \to \infty$. Smaller $L$ here is motivated by the smaller stationary variance of the cubic process for the choice of parameters above. For $M\Delta t = Const$ we use $\Delta t = 0.02, 0.01, 0.005, 0.002, 0.001$ and for $M\Delta t \to \infty$ we use $\Delta t = 0.01$.

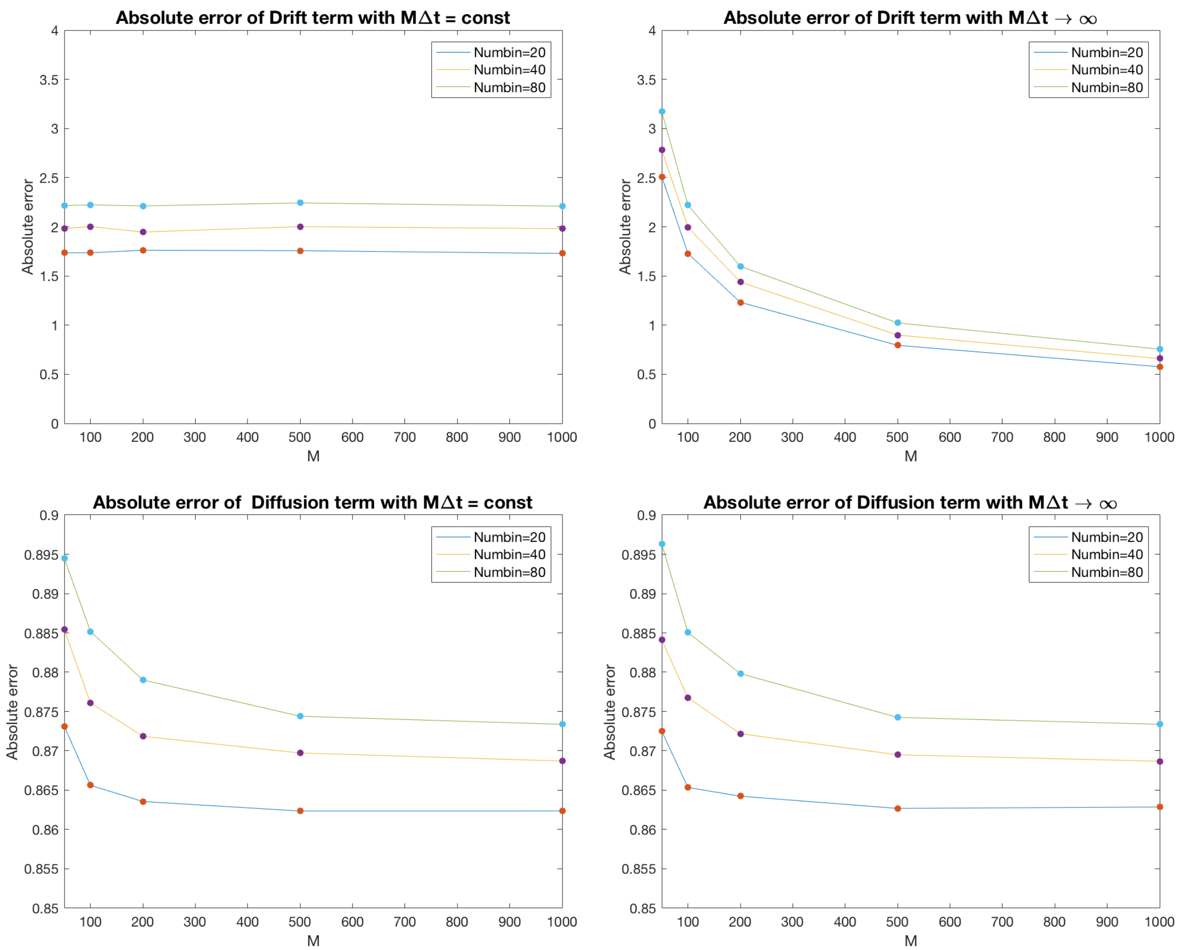



Figure 6: Absolute errors of Drift (top) and Diffusion (bottom) estimators for the cubic process with two different sampling regimes $M\Delta t = 500$ (left) and $M\Delta t \to \infty$ (right).

Figure 6 presents the absolute error of Drift and Diffusion estimators with two different sampling regimes $M\Delta t = Const$ and $M\Delta t \to \infty$ for cubic process. This figure is consistent with simulations for the OU process (cf., with Figure 3). In particular, we observe that the error behaves very differently for the Drift estimator in two regimes (no decay for $M\Delta t = Const$ vs. decay for $M\Delta t \to \infty$), while decay of error for the Diffusion estimator is approximately the same in the two regimes. We also observe that change in $\Delta x$ has a slightly more pronounced affect for the cubic process. However, for the Diffusion estimator the vertical scale for the error is rather small, and the error difference due to different $\Delta x$ is $O(10^{-3})$. In addition, the error for the Diffusion estimator appear to stabilize for $M = 500, 1000$, but is still quite large, which suggests the importance of the other terms in the expression for the error.

Larger number of observational points, $M$.

We also perform the simulations with the larger values of $M$ in (40). The numerical results are presented in Figure 7 and Table 2. The absolute error for the Drift estimator decays for the whole range of $M \in [50, 5000]$, while the absolute error for the Diffusion estimator stabilizes after $M \approx 500$. This suggests that the error term $(M\Delta t)^{-1}$ for the Drift estimator has a larger pre-constant compared to the term $M^{-1}$ for the Diffusion estimator.

Table 2: Absolute errors of Drift and Diffusion estimators for the cubic process for $M\Delta t \to \infty$ on $M \in [50, 5000]$ for three cases $NumBin = 20$, $NumBin = 40$, and $NumBin = 80$ where $\Delta x = 2L/NumBin$.

| Absolute error | | | | | | |
|---|---|---|---|---|---|---|
| | Drift estimator | | | Diffusion estimator | | |
| $M$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ | $\Delta x = 0.0125$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ | $\Delta x = 0.0125$ |
| 50 | 2.510 | 2.781 | 3.172 | 0.872 | 0.8841 | 0.8963 |
| 100 | 1.728 | 1.995 | 2.219 | 0.8653 | 0.8768 | 0.8850 |
| 200 | 1.232 | 1.439 | 1.598 | 0.8642 | 0.8722 | 0.8798 |
| 500 | 0.7944 | 0.8987 | 1.022 | 0.8627 | 0.8695 | 0.8742 |
| 1000 | 0.5764 | 0.6599 | 0.7555 | 0.8629 | 0.8687 | 0.8734 |
| 2000 | 0.4356 | 0.4924 | 0.5573 | 0.8627 | 0.8687 | 0.87245 |
| 5000 | 0.3165 | 0.3518 | 0.3888 | 0.8626 | 0.8688 | 0.8720 |

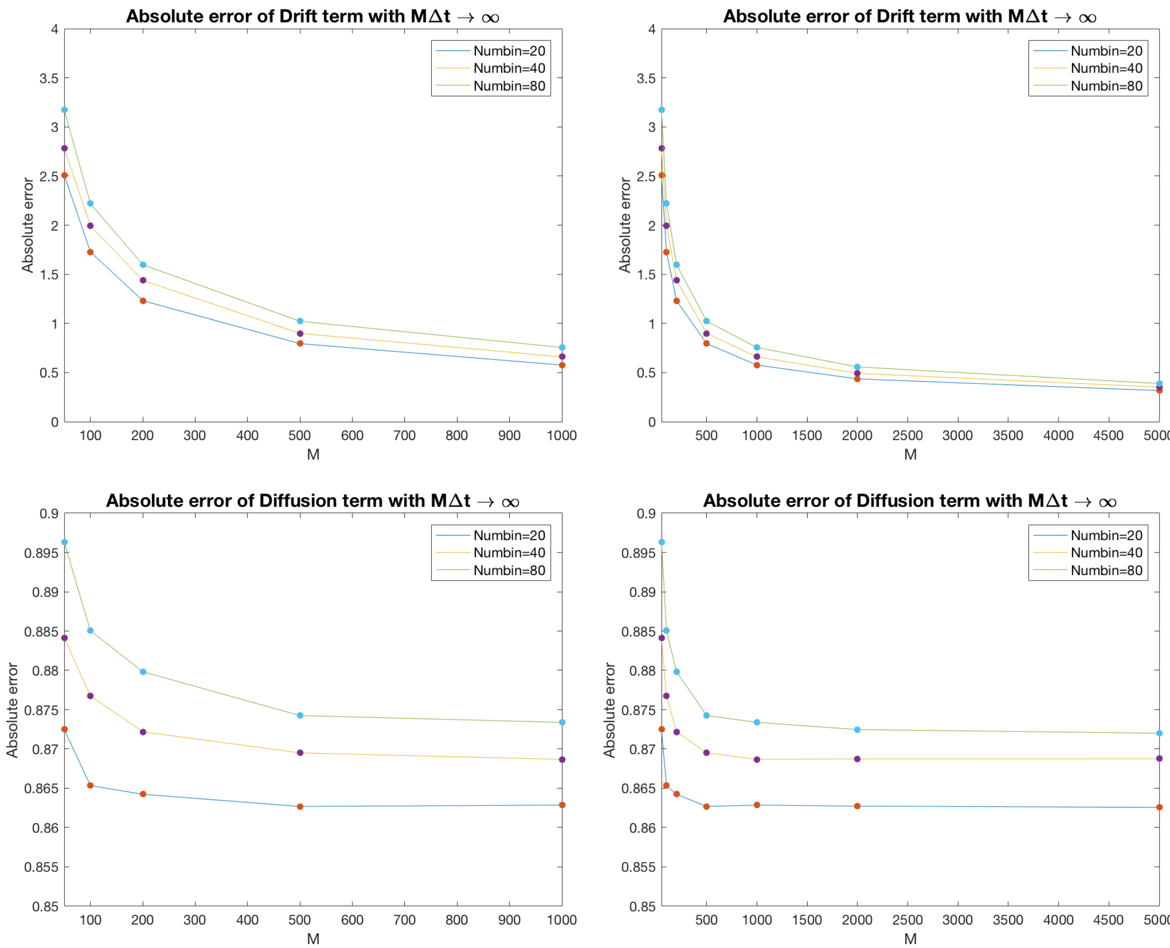



Figure 7: Absolute errors of Drift (top) and Diffusion (bottom) estimators for the cubic process for $M\Delta t \to \infty$ for two different ranges of $M \in [50, 1000]$ (left) and $M \in [50, 5000]$ (right).

<u>Smaller $\Delta x$ and $\Delta t$.</u>

Similar to the OU case, we consider the number of points, $M$, in (41) with $\Delta t = 0.002$ and $\Delta x = 2L/160$. The numerical results are presented in Figure 8. This picture is also consistent with the numerical results for the OU process depicted in Figure 5. **Drift estimator.** In particular, we observe that the absolute error for the Drift estimator *increases* approximately 3 times large for $M = 5000$ when $(\Delta t, \Delta x) = (0.002, 2L/160)$ vs. $(\Delta t, \Delta x) = (0.01, 2L/80)$ (cf., the left part of Figure 8 and the top-right part of Figure 7). We would like to point out that we expect that the results for the cubic process are more generic because it has a more general Itô-Taylor expansion.

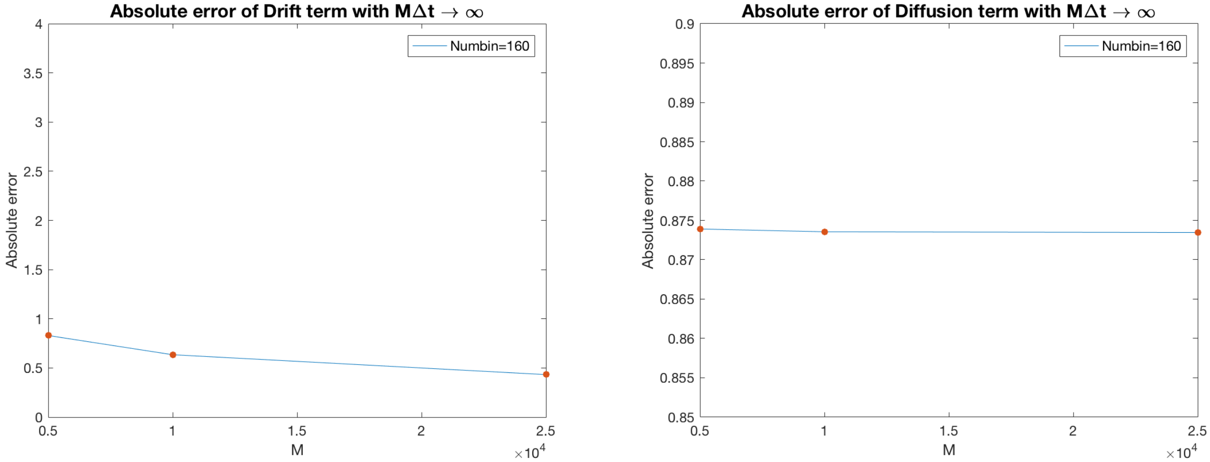



Figure 8: Absolute errors of the Drift and Diffusion term estimators for the cubic process with smaller $\Delta t = 0.002$ and $\Delta x = 2L/160$ and $M$ in (41).

### 3.3.3 Conclusions

In this section we analyzed numerically the absolute errors for the Drift and Diffusion estimators (defined in (36) and (37) respectively). These errors are practically equivalent to the $L^1$ errors. Although we do not have analytical estimates for the behaviors of these errors, we can obtain valuable guidelines for the role of the sub-sampling parameters $M$, $\Delta t$, and $\Delta x$.

The main emphasis of section 3.3 is on comparing and contrasting behavior of the absolute errors in the two sampling regimes $M\Delta t = Const$ and $M\Delta t \to \infty$. We analyzed the performance of these errors for two particular stochastic processes - the Ornstein-Uhlenbeck (OU) process (32) and the cubic process (33). The numerical results for these two processes are consistent with each other. Overall, our numerical simulations in section 3.3 suggest the following conclusions -

- there is a term $O((M\Delta t)^{-1})$ in the error for the Drift estimator,
- there is a term $O(M^{-1})$ in the error for the Diffusion estimator,
- there is a term with $\Delta t$ in the denominator in the error for the Drift estimator,
- refining $\Delta x$ makes a secondary effect on accuracy of estimators.

The numerical results and conclusions reached in this section are consistent with our analytical expressions for the MSE in (20) and (31). We will discuss practical guidelines for selecting estimation parameters in the section for the MSE for the Drift and Diffusion estimators.

### 3.4 Mean squared error for the Drift and Diffusion estimators

In this section, we use the MSE as a gauge for verifying the quality of Drift and Diffusion estimator with respect to the three parameters $M$, $\Delta t$ and $\Delta x$. The MSE is defined as (17) and (24) and is equivalent to the $L^2$ norm squared. We would like to remind that our analytical prediction for the asymptotic behaviors of the MSE was derived in (20) and (31) for the Drift and Diffusion estimators, respectively. Therefore, the goal of this section is to verify numerically the analytical expressions for the asymptotic behavior of the MSEs for the Drift and Diffusion estimators.

To compute the MSE numerically we need to perform Monte-Carlo simulations and compute many realizations of the sampled trajectories and, in turn, of the Drift and Diffusion estimators. Let us introduce the discrete analog of the MSE

$$MSE_{drift} = \frac{1}{MC}\sum_{j=1}^{MC}\left(\sum_k \left(\hat{A_k}^{(j)} - A(k)\right)^2 \Delta x\right), \tag{43}$$

$$MSE_{diff} = \frac{1}{MC}\sum_{j=1}^{MC}\left(\sum_k \left(\hat{B_k^2}^{(j)} - B^2(k)\right)^2 \Delta x\right), \tag{44}$$

where $MC$ is the number of Monte-Carlo Realizations, $\hat{A_k}^{(j)}$ and $\hat{B_k^2}^{(j)}$ are the Drift and Diffusion estimators computed for the $j$th Monte-Carlo Realizations, respectively, and $k$ represents the $k$-th bin (all bins are of size $\Delta x$).

#### 3.4.1 Ornstein-Uhlenbeck process

For the OU process, the Drift and Diffusion coefficients are given by

$$A(k) = -\gamma x_k, \qquad B^2(k) = \sigma^2,$$

with the parameters described in section 3.1.1. We consider the same sampling parameters as in section 3.3.1 and perform $MC = 500$ Monte-Carlo simulations in the estimation regimes $M\Delta t \to \infty$ and $M\Delta t = Const.$

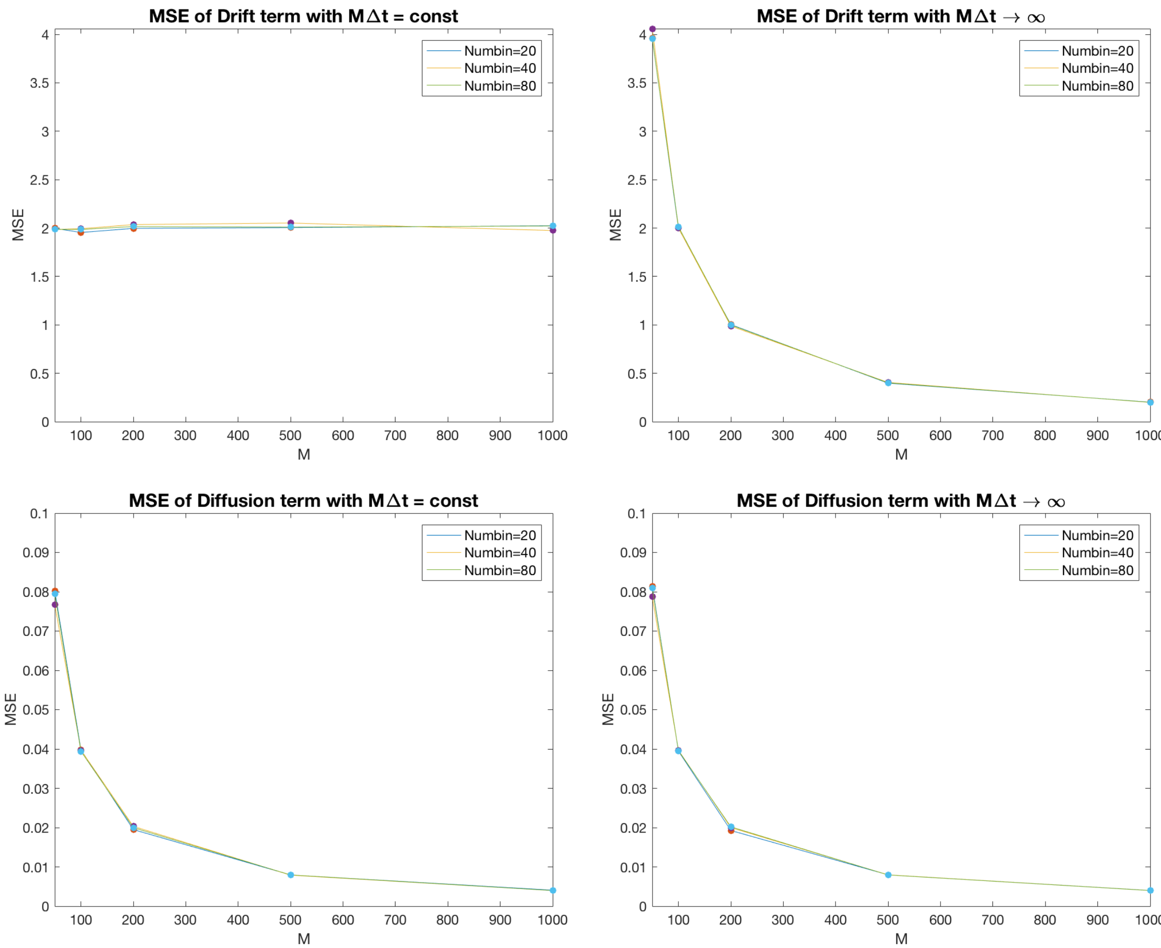



Figure 9: MSEs of the Drift (top) and Diffusion (bottom) estimators for the OU process with two different sampling regimes $M\Delta t = 500$ (left) and $M\Delta t \to \infty$ (right).

The numerical results for the asymptotic behaviors of the MSEs are presented in Figure 9. Overall, the behavior of MSEs for the Drift and Diffusion estimators is consistent with our analytical predictions in (20) and (31). In particular, the MSE for the Drift estimator in the regime $M\Delta t = Const$ does not decay as $M$ increases. We also observe that for $M\Delta t \to \infty$, the MSEs for both, the Drift and the Diffusion estimators do not stabilize and keep decreasing as $M \to 1000$. In addition, we also observe that decreasing (or increasing) the bin size $\Delta x$ in the range $\Delta x \in [2L/80, 2L/20]$ (with $L = 1$) does not have a visible affect on the accuracy of both, the Drift and Diffusion estimators. This suggests that $\Delta x$ can be chosen quite large in practice.

Larger number of observational points, $M$.

In order to estimate the relative importance of terms $(M\Delta t)^{-1}$ and $M^{-1}$ for the MSEs of the Drift and Diffusion estimator, respectively, we consider larger number of points in each bin, $M$, in (40).

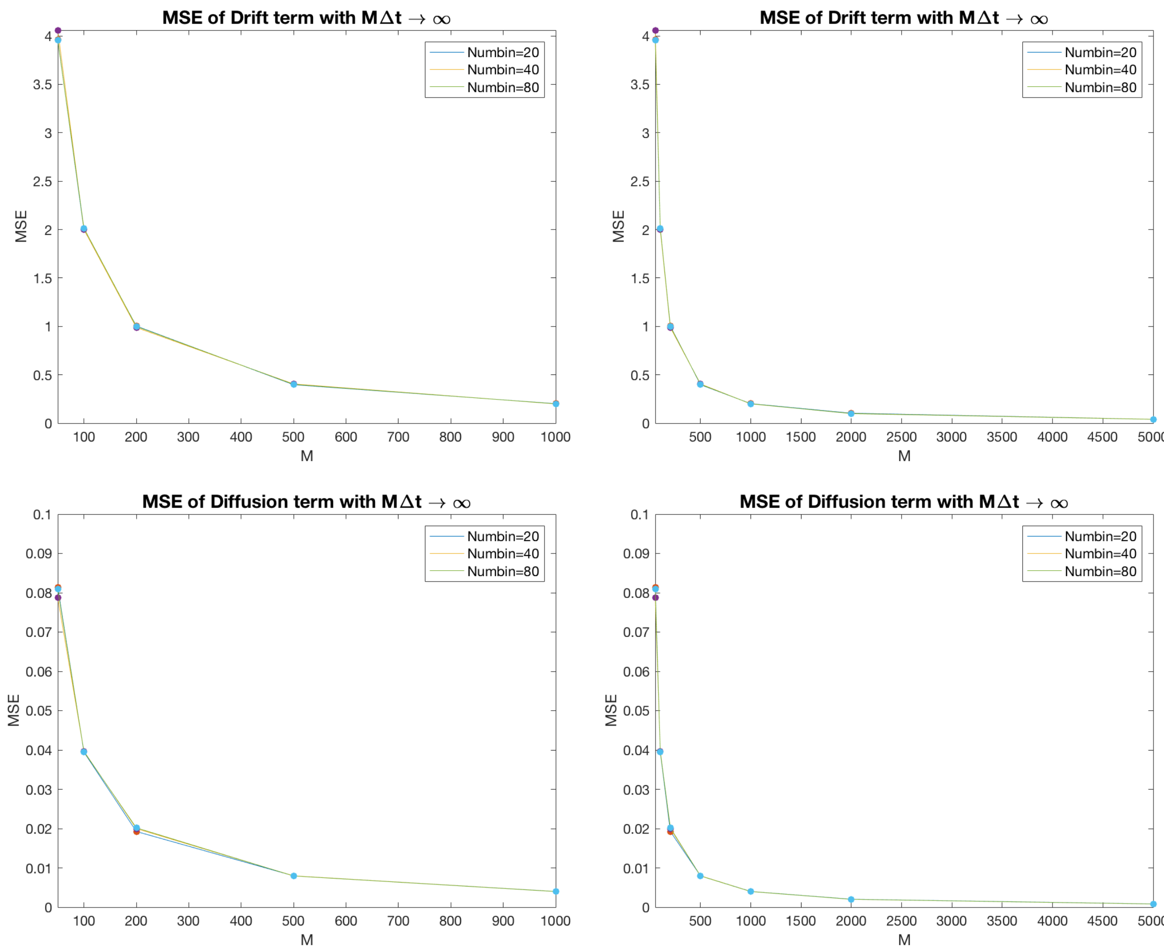



Figure 10: MSEs of Drift (top) and Diffusion (bottom) estimators for the OU process for $M\Delta t \to \infty$. for two different ranges of $M \in [50, 1000]$ (left) and $M \in [50, 5000]$ (right).

The results of these simulations are presented in Figure 10. First, MSE for the Drift estimator is much larger than the MSE for the Diffusion estimator. For $M = 1000$ the Diffusion estimator appears to be quite accurate, while the Drift estimator has significant errors. We also observe that MSEs for both the Drift and the Diffusion estimators do not stabilize in the range $M \in [50, 5000]$, which suggests the relative importance of terms $(M\Delta t)^{-1}$ and $M^{-1}$ for the MSEs of the Drift and Diffusion estimator, respectively, for the whole range of $M$ considered here.

Numerical results are presented in Figure 10 and Table 3.

Table 3: MSEs of Drift and Diffusion estimators for the OU process for $M\Delta t \to \infty$ on $M \in [50, 5000]$ for three cases $NumBin = 20$, $NumBin = 40$ and $NumBin = 80$ where $\Delta x = 2L/NumBin$.

| MSE | | | | | | |
|---|---|---|---|---|---|---|
| | Drift estimator | | | Diffusion estimator | | |
| $M$ | $\Delta x = 0.1$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ | $\Delta x = 0.1$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ |
| 50 | 3.963 | 4.058 | 3.957 | 0.08138 | 0.07871 | 0.08092 |
| 100 | 1.998 | 1.999 | 2.013 | 0.03945 | 0.03967 | 0.03956 |
| 200 | 1.003 | 0.9880 | 0.999 | 0.01926 | 0.01999 | 0.02019 |
| 500 | 0.3977 | 0.4052 | 0.4011 | 0.007983 | 0.007982 | 0.0079477 |
| 1000 | 0.2018 | 0.1995 | 0.2006 | 0.004009 | 0.003994 | 0.003982 |
| 2000 | 0.1035 | 0.1003 | 0.0990 | 0.002032 | 0.002005 | 0.002009 |
| 5000 | 0.04046 | 0.04061 | 0.04010 | 0.0008397 | 0.00082198 | 0.0008161 |

<u>Smaller $\Delta x$ and $\Delta t$.</u>

We also consider the much larger values of $M$ in (41) with $\Delta t = 0.002$.

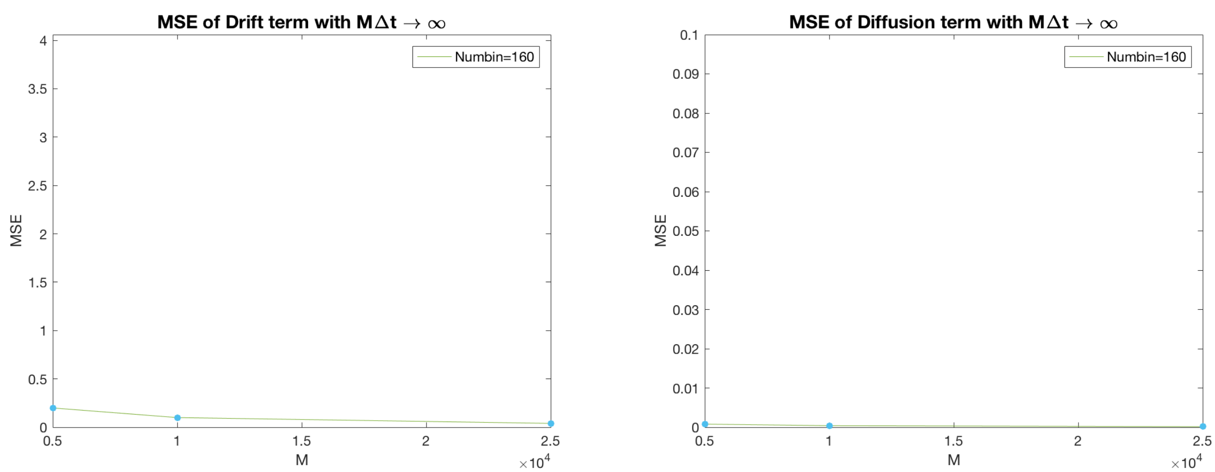



Figure 11: MSEs of the Drift and Diffusion term estimators for the OU process with smaller $\Delta t = 0.002$ and $\Delta x = 2L/160$ and $M$ in (41).

The numerical results are presented in Figure 11. We observe that the behaviors of MSEs for the Drift and Diffusion estimators are consistent with (20) and (31). In particular, the MSE for the Drift estimator increases for $M = 5000$ and smaller $\Delta t$ (cf., the left part of Figure 11 for $(M, \Delta t) = (5000, 0.002)$ and the top-right part of Figure 10 for $(M, \Delta t) = (5000, 0.01)$).

We also observe that MSEs for both, Drift and Diffusion estimators keep decaying in the range $M \in [5000, 25000]$. This indicates the significance of terms $(M\Delta t)^{-1}$ and $M^{-1}$ for the MSEs of the Drift and Diffusion estimator, respectively. However, the MSE for the Diffusion estimator is much smaller than the MSE for the Drift estimator. This implies that the Drift estimator is less accurate (especially for small $\Delta t$) than the Diffusion estimator and estimation of the Drift term can be quite computationally expensive for small $\Delta t$. This suggests that one can possibly use two different sampling time-steps for the Drift and Diffusion estimators - a larger sampling time-step to compute the Drift estimator and a smaller sampling time-step to compute the Diffusion estimator. We'll also address this issue in subsequent sections.

### 3.4.2 Cubic process

For the cubic process (33) the Drift and Diffusion coefficients are given by

$$A(k) = -\gamma x_k^3, \qquad B^2(k) = (\sigma_1 + \sigma_2 x_k)^2.$$

We consider the parameter values as in section 3.1.2. We choose the same computational and sampling parameters as in section 3.3.2. We consider values of $M$ in (38) and we also operate three sets of runs with $\Delta x$ in (42) in the regimes $M\Delta t \to \infty$ and $M\Delta t = Const$. Similar to other sections we perform $MC = 500$ Monte-Carlo simulations.

Figure 12 illustrates the behaviors of MSEs for the Drift and Diffusion estimators in two different sampling regimes $M\Delta t = Const$ and $M\Delta t \to \infty$. Our results for the behavior of MSEs for the cubic process are consistent with previous results for the absolute errors in sections 3.3.1 (OU process) and 3.3.2 (cubic process) and MSEs for the OU process in section 3.4.1. The results are also consistent with analytical predictions in (20) and (31). Neither the MSE for the Drift estimator nor the MSE for the Diffusion estimator stabilizes in the range $M \in [50, 1000]$ in the regime $M\Delta t \to \infty$. This suggests that terms $(M\Delta t)^{-1}$ and $M^{-1}$ in the MSEs of the Drift and Diffusion estimator, respectively, are quite significant for this range of $M$. In addition, the smaller

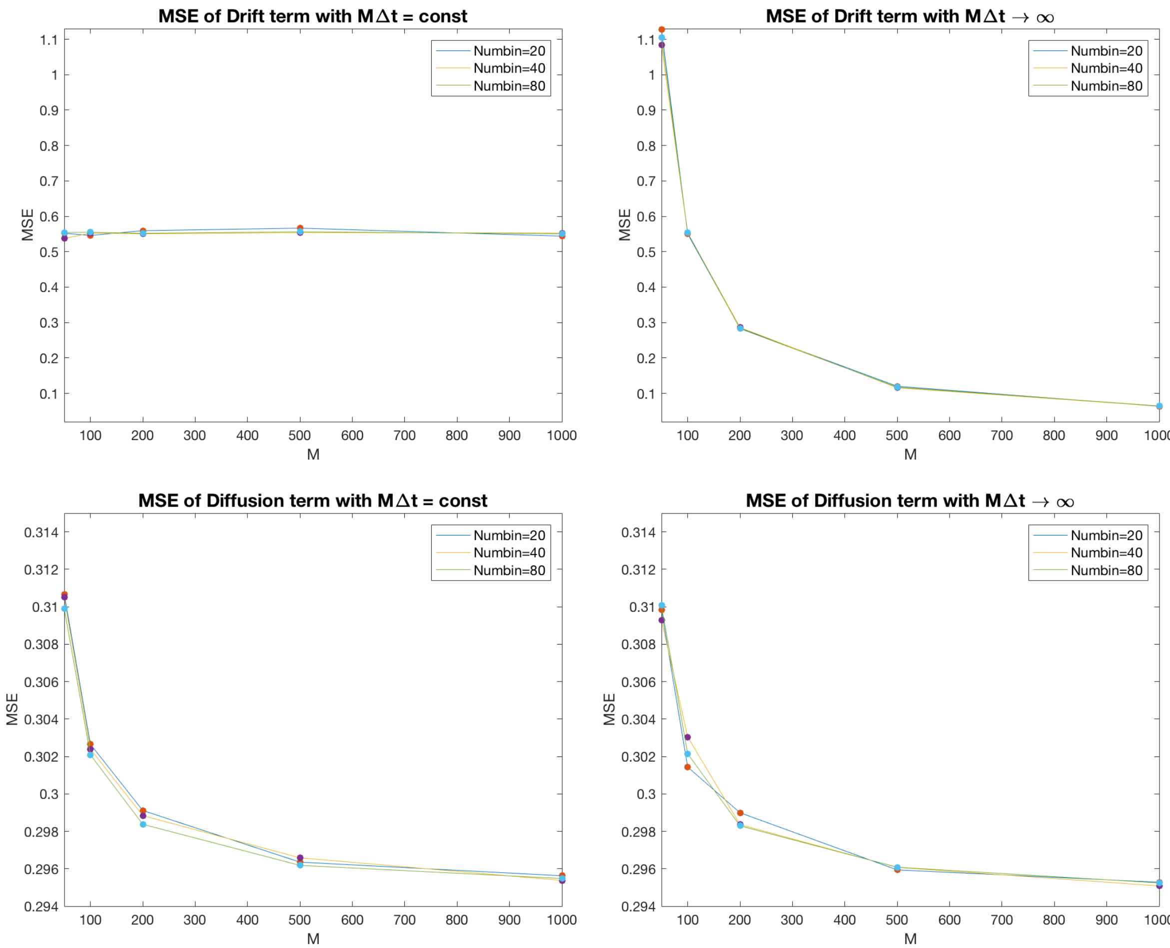



Figure 12: MSEs of Drift (top) and Diffusion (bottom) estimators for the cubic process with two different sampling regimes $M\Delta t = 500$ (left) and $M\Delta t \to \infty$ (right).

values of $\Delta x$ do not have any visible effect on the accuracy of the Drift estimator, while changes in $\Delta x$ have only slight effect for the accuracy of the Diffusion estimator. Also, similar to the simulations for the OU process, the MSE for the Drift estimator is larger (approximately twice in this case) than the MSE for the Diffusion estimator.

Larger number of observational points, $M$.

To investigate numerically the range of the significant values of terms $(M\Delta t)^{-1}$ and $M^{-1}$ in the MSEs for the Drift and Diffusion estimators, respectively, we consider number of the sampling points in each bin, $M$, in (40).

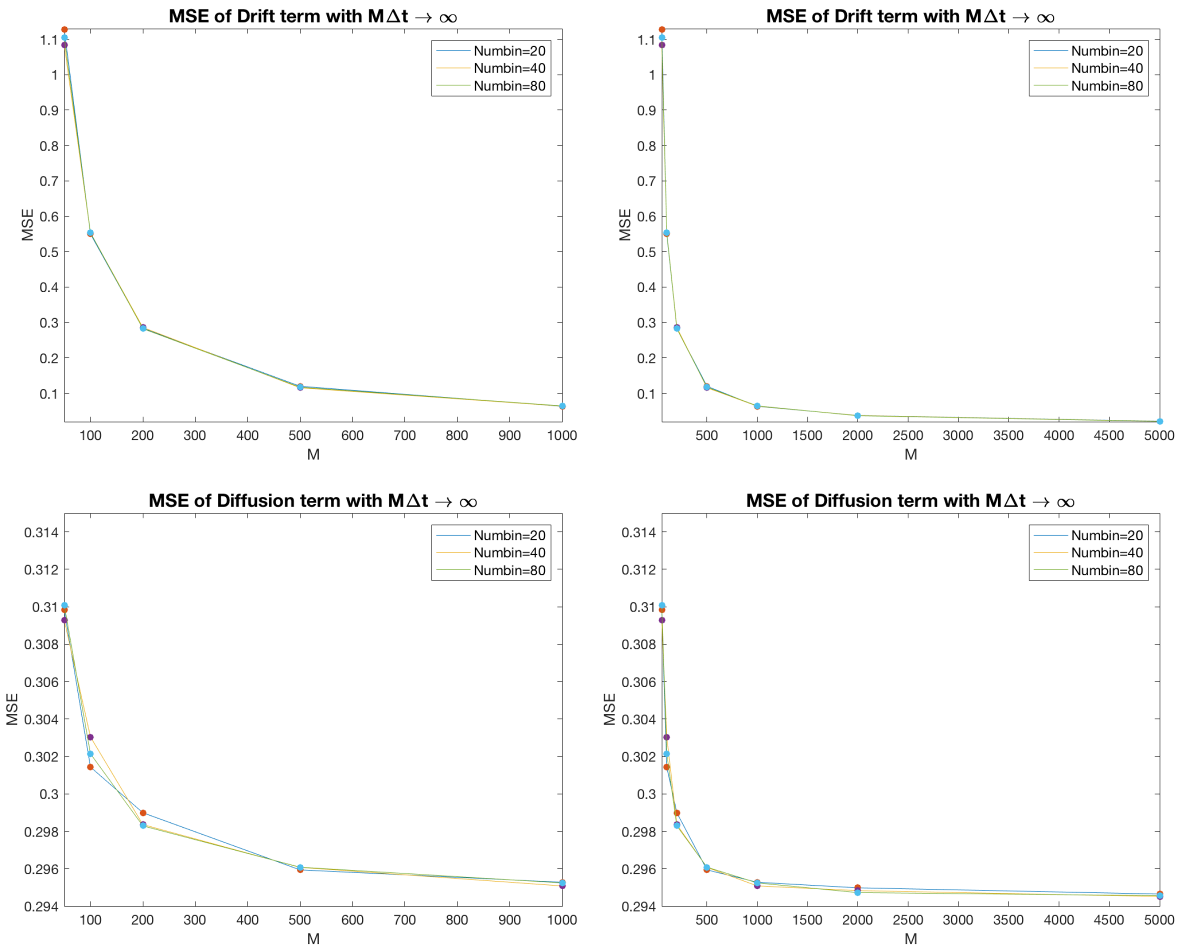



Figure 13: MSEs of Drift (top) and Diffusion (bottom) estimators for the cubic process for $M\Delta t \to \infty$. for two different ranges of $M \in [50, 1000]$ (left) and $M \in [50, 5000]$ (right).

Figure 13 shows the comparison of MSEs for the Drift and Diffusion estimators for $M \in [50, 1000]$ (left part) and $M \in [50, 5000]$ (right part) in the regime $M\Delta t \to \infty$. We observe that the MSE for the Diffusion estimator stabilizes in the range $M \in [1000, 5000]$, while the MSE for the Drift estimator is decaying for the whole range $M \in [50, 5000]$. This suggests that for the Diffusion estimator the other terms in the (31) become significant in the range $M \in [1000, 5000]$.

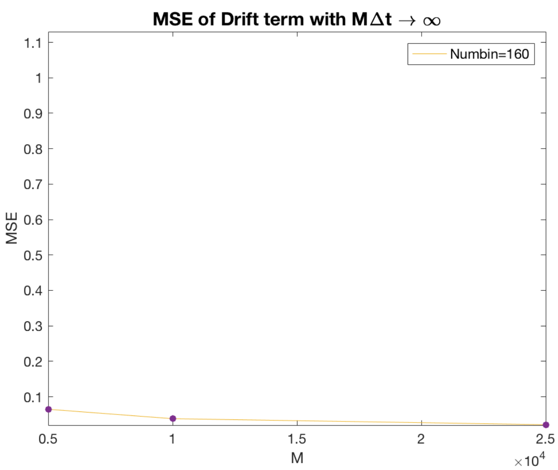
The numerical results are presented in Figure 13 and Table 4.

Smaller $\Delta x$ and $\Delta t$.

Table 4: MSEs of Drift and Diffusion estimators for the cubic process for $M\Delta t \to \infty$ on $M \in [50, 5000]$ for three cases $Numbin = 20$, $Numbin = 40$ and $Numbin = 80$ respectively.

| MSE | | | | | | |
|---|---|---|---|---|---|---|
| | Drift estimator | | | Diffusion estimator | | |
| $M$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ | $\Delta x = 0.0125$ | $\Delta x = 0.05$ | $\Delta x = 0.025$ | $\Delta x = 0.0125$ |
| 50 | 1.127 | 1.084 | 1.104 | 0.3098 | 0.3093 | 0.3101 |
| 100 | 0.5504 | 0.5534 | 0.5536 | 0.3014 | 0.3030 | 0.3021 |
| 200 | 0.2838 | 0.2862 | 0.2827 | 0.2990 | 0.2984 | 0.2983 |
| 500 | 0.1201 | 0.1158 | 0.1174 | 0.2959 | 0.2961 | 0.2961 |
| 1000 | 0.06335 | 0.06392 | 0.06461 | 0.2953 | 0.2951 | 0.2952 |
| 2000 | 0.03770 | 0.03746 | 0.03689 | 0.2950 | 0.2948 | 0.2947 |
| 5000 | 0.02139 | 0.02109 | 0.02085 | 0.2946 | 0.2945 | 0.2946 |

Here we consider the much larger values of $M$ in(41) with $\Delta t = 0.002$



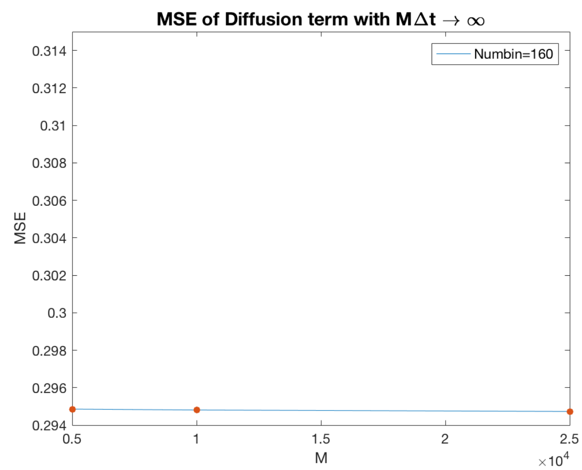



Figure 14: MSEs of the Drift and Diffusion term estimators for the cubic process with smaller $\Delta t = 0.002$ and $\Delta x = 2L/160$ and $M$ in (41).

Figure 14 depicts the MSEs for the Drift and Diffusion estimators in this regime. The MSE of the Diffusion estimator is not affected by changes in $M$. Therefore, the Diffusion estimator does not benefit from increasing the number of the sampling points, $M$, beyond $M = 1000$ (cf., the bottom-right part of Figure 13 and the right part of Figure 14).

### 3.4.3 Conclusions

In sections 3.4.1 and 3.4.2 we considered the asymptotic behaviors of the MSEs for the Drift and Diffusion estimators for two particular examples - the Ornstein-Uhlenbeck process and the cubic process. We compared and contrasted behavior of the MSEs for these two processes in the two sampling regimes $M\Delta t = Const$ and $M\Delta t \to \infty$. We also varied the size of the bin, $\Delta x$. The main conclusions reached by considering the numerical simulations are

- our numerical simulations agree well with our analytical predictions in (20) and (31),
- numerical simulations in this section agree with numerical results for the absolute error in section 3.3.1 and 3.3.2
- our numerical simulations confirm that there are terms $O((M\Delta t)^{-1})$ and $O(M^{-1})$ in the MSEs for the Drift and Diffusion estimators, respectively,
- the MSE for the Drift estimator appear to be large than the MSE for the Diffusion estimator for the same choice of computational parameters $M$, $\Delta t$, and $\Delta x$,
- conclusions reached in this section with conclusions for the absolute error discussed in section 3.3.3,
- it might be computationally beneficial to choose different sampling regimes for the Drift and Diffusion estimators.

We elaborate a little bit more about important practical guidelines on selecting the sampling and computational parameters $M$, $\Delta t$, and $\Delta x$ to reduce computational complexity while maintaining accuracy of estimators. In particular, it might be beneficial to select different sampling regimes for computing the Drift and Diffusion estimators. Here are the practical guidelines motivated by our numerical simulations -

- very small sampling time-step $\Delta t$ is very likely to negatively impact the accuracy of the Drift estimator; therefore, $\Delta t$ should be selected to be relatively large in the computation of the Drift estimator,

- sampling time-step, $\Delta t$, does not seem to severely affect the accuracy of the Diffusion estimator; therefore, $\Delta t$ can be selected to be quite small for computing the Diffusion estimator which can yield in necessity to process relatively short time-series of observations,

- bin size, $\Delta x$, in the range considered here does not seem to have a significant impact on accuracy of both, the Drift and the Diffusion estimators; therefore $\Delta x \approx L/10$ (where $L = StdDev\{X_t\}$) appears to be a good practical guideline for selecting the bin size.

## 3.5 Regression for the Drift and Diffusion terms

In the previous sections we investigated how Drift and Diffusion estimators perform for each bin. Our analysis applies to each bin separately and numerical simulations take the spacial distribution of errors into account in a rather simple form (either by taking a max or averaging). However, it is a very natural question to estimate the functional form of the Drift and Diffusion terms once we compute their estimators for each bin. It is quite beneficial to know the functional form of the Drift and Diffusion coefficients explicitly since it is possible to carry out a detailed analysis of the model (e.g., stability). Therefore, in this section we utilize regression techniques to estimate the functional form of the Drift and Diffusion coefficients from the numerical values obtained by computing the corresponding estimators. To this end, we assume a particular functional form for the Drift and Diffusion terms with estimate coefficients. We would like to point out that it is rather difficult to obtain the analytical error estimates for those coefficients because it is not clear how errors in the Drift and Diffusion estimator would "propagate" through the regression procedure. Thus, we concentrate here on some numerical results.

In this section, we compare the true Drift and Diffusion coefficients with computational results using statistical techniques. We test several approaches known to work well in practice – Polynomial fit, Lasso, and Ridge regression. We carry out th eregression estimation for both the Ornstein-Uhlenbeck process and the cubic process. To quantify the regression error we introduce the Root Mean Squared Error (RMSE)

$$RMSE = \left( \sum_{k=1}^{NB} \big(predict(x_k) - true(x_k)\big)^2 \right)^{1/2}, \tag{45}$$

which is a standard measure of accuracy in regression.

Since the bin size is not a dominant factor for accomplishing the accuracy of the Drift and Diffusion estimators as discussed in section 3.3 and 3.4, we use the numerical values for the Drift and Diffusion estimators obtained with the number of bins $NumBin = 40$. We would like to point out that the total number of bins is twice that, i.e., the total number of bins is $NB = 80 = 2NumBin$

since we have to take estimation for both, the positive and negative values of $x_k$. This corresponds to using numerical results for the Drift and Diffusion estimation with $\Delta x = 0.025$ for the OU process and with $\Delta x = 0.0125$ for the cubic process, respectively. We select few particular values of $M$ and $\Delta x$ from sets of simulations described in the previous section. In particular, we present the results for $(M, \Delta t) = (1000, 0.001)$ and $(M, \Delta t) = (1000, 0.01)$.

### 3.5.1 Polynomial fit

In this section we perform polynomial regression for the Drift and Diffusion coefficients. In particular, we assume a polynomial functional form

$$f_l(\mathbf{a}, x) = a_l x^l + \ldots + a_2 x^2 + a_1 x + a_0 \tag{46}$$

with vector $\mathbf{a} = (a_l, \ldots, a_0)$ (please note that coefficients $a_l$ are ordered backwards, but the numbering corresponds to the power of $x$ for each term) and optimize the norm

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \sum_k \Big( f_l(\mathbf{a}, x_k) - \hat{F}(x_k) \Big)^2 \tag{47}$$

where $\hat{F}(x_k)$ is either the Drift (i.e., $\hat{F}(x_k) \equiv \hat{A}(x_k)$) or the Diffusion (i.e., $\hat{F}(x_k) \equiv \hat{B}^2(x_k)$) estimator and $x_k$ is the center of the corresponding bin. The RMSE is then computed by substituting $f_l(\hat{\mathbf{a}}, x_k)$ as the predicted value $\mathit{predict}(x_k)$.

It is known that polynomial regression is likely to produce very oscillatory results if we take a (relatively) large $l$ (highest power) in the definition of $f_l(\mathbf{a}, x)$ in (46). Therefore, plain polynomial regression is rarely used in practice. Thus, we just want to obtain some benchmark results in this section. Therefore, we choose $l$ to be the lowest possible power for the Drift and Diffusion coefficients for each model (OU and cubic processes).

## OU process.

Here we choose $l = 1$ for the Drift and $l = 0$ for Diffusion. The numerical results are presented in

Figure 15 and Table 5. We can see that the smaller $\Delta t$ makes a big difference for the accurate

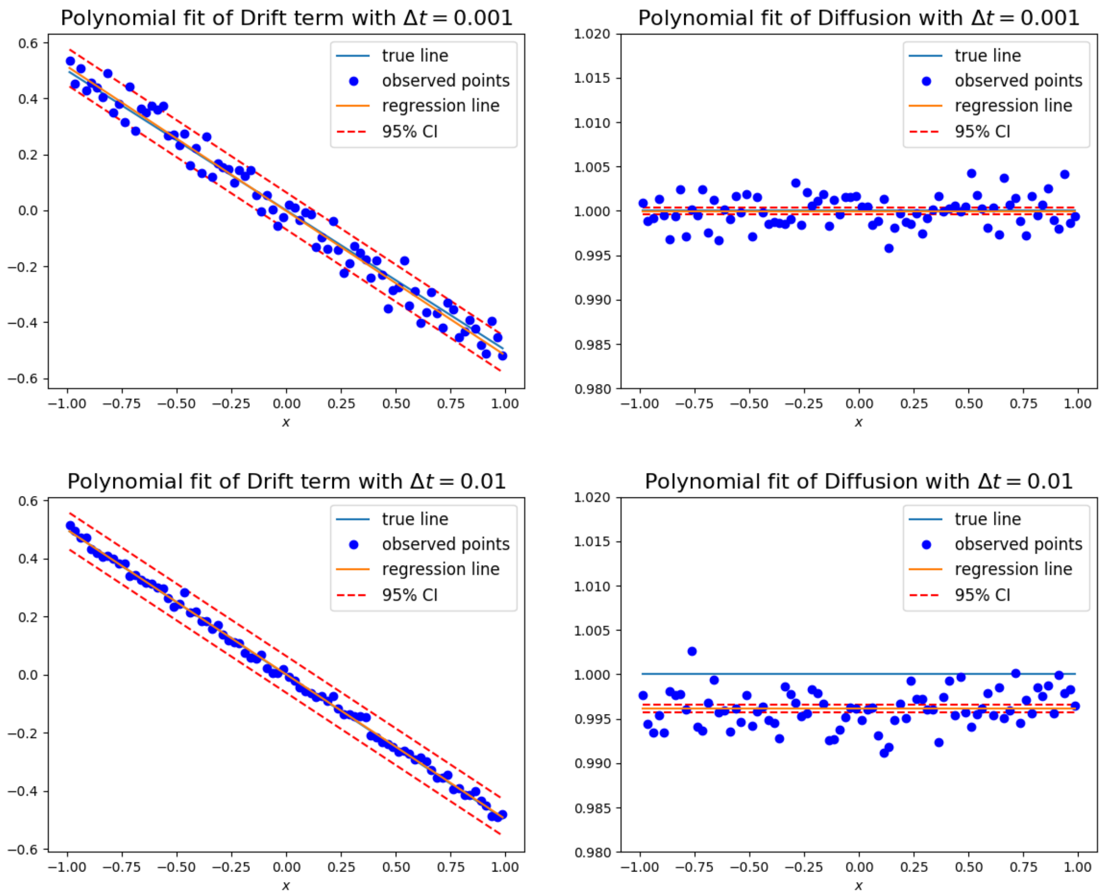



Figure 15: Polynomial regression fit for the Drift (left part) and Diffusion (right part) terms of the OU process with $(M, \Delta t) = (1000, 0.001)$ (top part) and $(M, \Delta t) = (1000, 0.01)$ (bottom part).

estimation of the Drift term (cf., the two sub-plots in the left part of Figure 15). In particular, the estimated numerical values of $\hat{A}(x)$ are much more dispersed for larger $\Delta t = 0.01$. This is also manifested in a larger RMSE for the Drift term for $\Delta t = 0.01$. However, Table 5 illustrates that the Drift term estimator is more accurate for $\Delta t = 0.01$. In contrast to the Drift estimation, the estimation of the Diffusion term is not considerably affected by increasing $\Delta t$. In particular, the estimated numerical values of $\hat{B}^2(x)$ do not appear to be dispersed very differently for $\Delta t = 0.001$ vs. $\Delta t = 0.01$. The estimated value changes slightly for $\Delta t = 0.01$, but Table 5 indicates that the RMSE for the Diffusion term is not affected much by a larger $\Delta t$. Finally, Table 5 demonstrates

Table 5: Polynomial regression fit results for the OU process.

| $(M, \Delta t)$ | | Drift coef. | Diffusion coef. |
|---|---|---|---|
| $(1000, 0.001)$ | Estimated | $-0.5171x - 0.0022$ | 1 |
| | RMSE | 0.4136 | 0.0158 |
| $(1000, 0.01)$ | Estimated | $-0.4984x + 0.0015$ | 0.9961 |
| | RMSE | 0.1300 | 0.0190 |

that the Diffusion term estimator performs better than Drift estimator, since the RMSE is smaller for Diffusion estimator.

### Cubic process.

We select $l = 3$ and $l = 1$ for the Drift and Diffusion estimation, respectively. The results are presented in Figure 16 and Table 6. Here we can see that the Drift term is estimated much better

Table 6: Polynomial regression fit results for the cubic process.

| $(M, \Delta t)$ | | Drift coef. | Diffusion coef. |
|---|---|---|---|
| $(1000, 0.001)$ | Estimated | $-0.57x^3 + 0.038x^2 - 0.055x - 0.0046$ | $0.499x^2 + 0.999x + 0.5$ |
| | RMSE | 0.2479 | 0 |
| $(1000, 0.01)$ | Estimated | $-0.953x^3 - 0.0106x^2 - 0.012 - 0.0002$ | $0.485x^2 + 0.997x + 0.501$ |
| | RMSE | 0.0792 | 0.0114 |

with $\Delta t = 0.01$ (the bottom part of Figure 16). This is also supported by a much smaller RMSE for the Drift term for $\Delta t = 0.01$ (Table 6). The Diffusion term is estimated comparably accurately for $\Delta t = 0.01$ and $\Delta t = 0.001$. The RMSE for the Diffusion term is only slightly bigger for $\Delta t = 0.01$ compared to $\Delta t = 0.001$.

### 3.5.2 Estimation of the Diffusion coefficient for the OU process for different $\Delta t$

In this section we discuss the behavior of the Diffusion term estimator for $(M, \Delta t) = (1000, 0.01)$ vs. $(M, \Delta t) = (1000, 0.001)$. This behavior is depicted in the right part of Figure 15 and Table 5. In particular, the estimator is less accurate for $\Delta t = 0.01$. In particular, the estimator for the constant Diffusion shifts slightly ($\hat{B}^2(x) = 1$ for $\Delta t = 0.001$ vs. $\hat{B}^2(x) = 0.996$ for $\Delta t = 0.01$) which results in a slightly higher RMSE for $\Delta t = 0.01$. Similar to the previous section we use $NB = 80$ (same as $\Delta x = 0.025$).

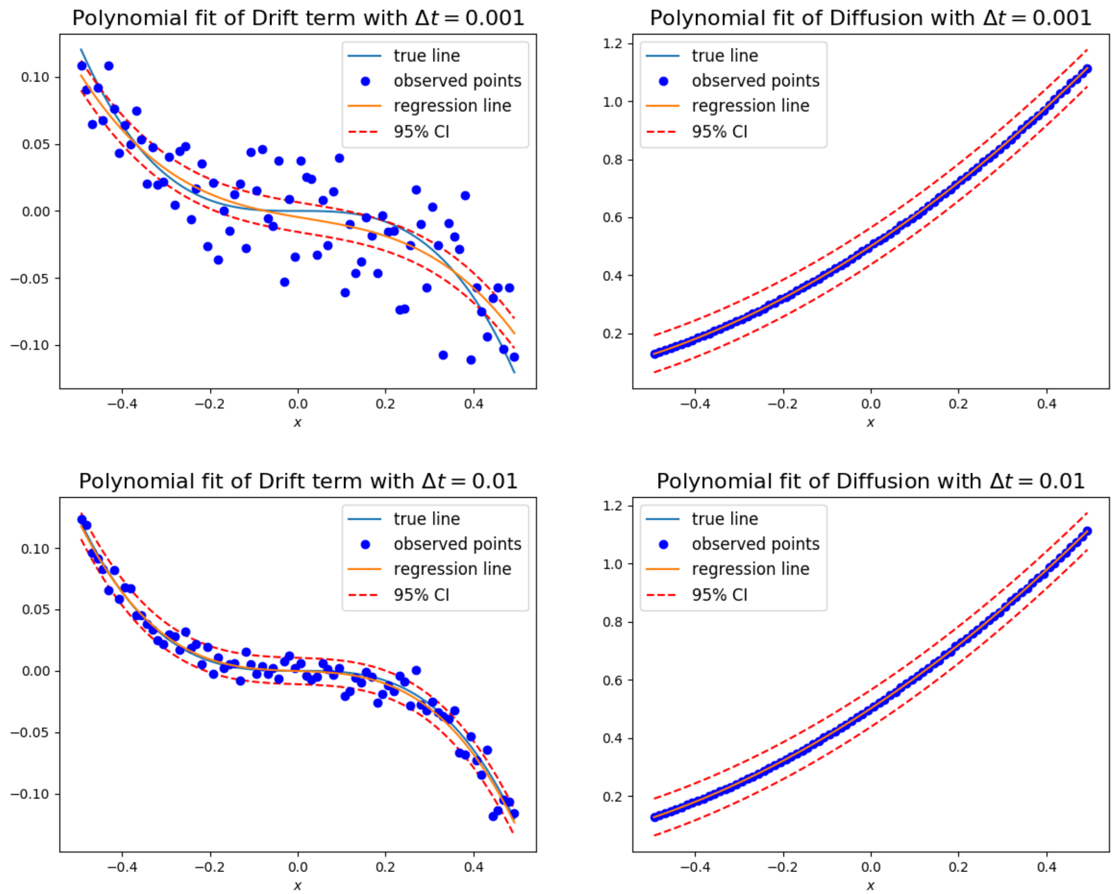



Figure 16: Polynomial regression fit for the Drift (left part) and Diffusion (right part) terms of the cubic process with $(M, \Delta t) = (1000, 0.001)$ (top part) and $(M, \Delta t) = (1000, 0.01)$ (bottom part).

Since the Diffusion is just a constant, it is relatively easy to analyze the behavior of the error for the regression for the Diffusion term. We would like to remind that we use $l = 0$ as the highest-order power in the polynomial for the Diffusion term. We define the bias

$$Bias_{drift} = |a_1 - \gamma|, \qquad Bias_{diff} = |a_0 - \sigma|, \tag{48}$$

where $a_1$ is the coefficient for the linear term in the regression for the Drift and $a_0$ is the coefficient for the free (constant) term in the regression for Diffusion. The particular values of parameters are

$\gamma = 1/2$ and $\sigma = 1$. We vary $\Delta t$ over

$$\Delta t = 0.02, 0.01, 0.005, 0.002, 0.001 \tag{49}$$

and compute bias for the Drift and Diffusion estimators for these five value of $\Delta t$. Figure 17

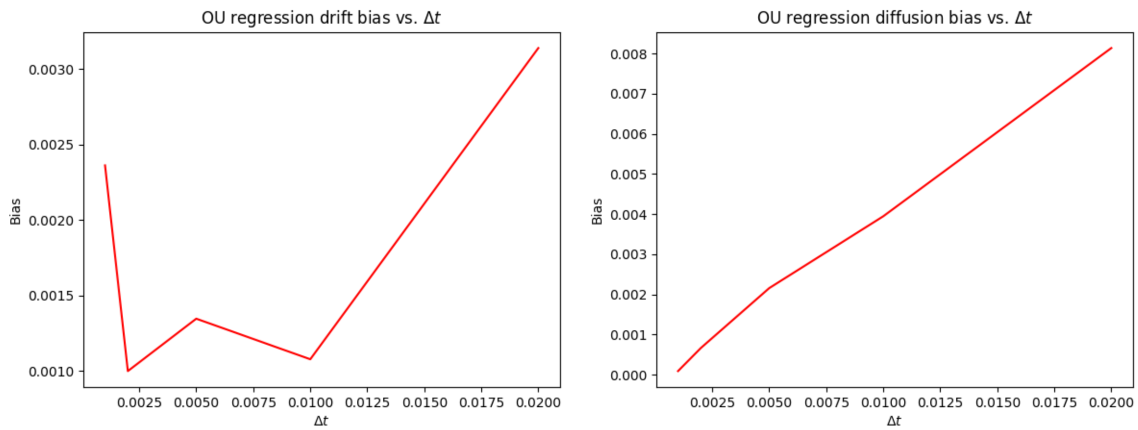



Figure 17: Regression Bias of the Drift (left part) and Diffusion (right part) terms in (48) computed from simulations of the OU process with $M = 1000$ and $\Delta t$ in (49).

shows that the bias for the Diffusion decreases approximately linearly with $\Delta t$ and $Bias \to 0$ as $\Delta t \to 0$. This demonstrates that the accuracy of the regression for the Diffusion estimator depends considerably on $\Delta t$ and it is important to have a small $\Delta t$ if a high accuracy of the Diffusion estimation is the objective. On the other hand, the numerical errors for the Diffusion increase only slightly as $\Delta t$ increases and larger $\Delta t = 0.02$ considered here the results in relatively small errors for the Diffusion estimation. Figure 17 demonstrates that the bias for the Drift term has a U-shape and increases for very small $\Delta t$. This illustrates that small values of $\Delta t$ should be avoided for estimating the Drift term. The simulations in this section are consistent with our numerical analysis for the absolute error and MSE in sections 3.3 and 3.4, respectively and regression results in the previous section.

### 3.5.3 Polynomial, Lasso, and Ridge estimation for the cubic model

In the section 3.5.1 we observed that the polynomial regression works quite well to estimate the shape of the Drift and Diffusion terms when the highest power of the fitted polynomial is low (i.e., $l$ is small in (46)). It is well-known that polynomial regression becomes unstable and produces very large coefficients for the higher values of $l$. Therefore, we utilize Lasso and Ridge regression. We consider $l = 7$ (i.e., we try to fit polynomials of 7-th order) and compare the results for the polynomial, Lasso, and Ridge regressions.

Here we concentrate on the numerical results for the cubic process since for the OU process it is very easy to estimate visually the functional form of the Drift and Diffusion terms. The results for the cubic process are more generic since it is not so easy to guess the functional form of the Drift and Diffusion terms for this model. We use the numerical results for estimating $\hat{A}(x_k)$ and $\hat{B}^2(x_k)$ with $(M, \Delta t) = (1000, 0.01)$ since we saw in the previous section that $\Delta t = 0.01$ results in considerably more accurate Drift estimation. Similar to the previous section we utilize $NB = 80$ as the total number of bins which corresponds to $\Delta x = 0.025$.

#### Polynomial Regression.

Here we use the same optimization function (45) as in section 3.5.1 with $l = 7$. We obtained the following fits for the polynomial regression of degree $l = 7$ for the Drift and Diffusion terms

$$\hat{\mathbf{a}}_{drift} = (19.17, 4.02, -8.49, -1.38, 0.134, 0.106, -0.0474, -0.00158), \qquad \text{RMSE} \approx 0.077,$$

$$\hat{\mathbf{a}}_{diff} = (1.831, 0.366, -0.529, -0.0705, 0.00392, 0.4833, 1.0011, 0.5012), \qquad \text{RMSE} \approx 0.010227.$$

We would like to remind that first coefficients correspond to power $x^l$ (see definition (46)). Fitting higher order polynomial with $q = 7$ slightly reduces the RMSE for both, Drift and Diffusion (cf., with results for $(M, \Delta t) = (1000, 0.01)$ in Table 6), but the reduction is not really significant. On the other hand, polynomial fit with degree $q = 7$ produces results which are completely misleading and cannot be used for predicting the Drift and Diffusion outside of the range $[-L, L]$. Therefore, if one wants to use the estimated model for numerical or analytical prediction, it is very likely that numerical results using the fitted model obtained by polynomial regression would not be accurate.

**Lasso Regression.**

In this section, we utilize Lasso to shrink the regression coefficients by adding a $L^1$ penalty term. The optimization problem becomes

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{NB} \sum_{k=1}^{NB} \Big( f_l(\mathbf{a}, x_k) - \hat{F}(x_k) \Big)^2 + \alpha \sum_{i=0}^{l} |a_i| \tag{50}$$

where $\hat{F}(x_k)$ are the numerical computed Drift and Diffusion coefficients $\hat{A}(x_k)$ and $\hat{B}^2(x_k)$. Here an empirical parameter $\alpha$ balances the relative strength of the polynomial regression term and penalty term. We would like to point out that $1/NB$ is just a normalization factor for the sum and can be “absorbed” in the parameter $\alpha$. Results for the Lasso regression for the Drift and Diffusion estimators computed with $\Delta t = 0.01$ are presented below.

$$\hat{\mathbf{a}}_{drift} = (0, 0, -0.26, 0, -0.866, 0, -0.014, -0.001),$$

$$\hat{\mathbf{a}}_{diff} = (0.0895, 0.0095, 0, 0.0234, -0.0223, 0.475, 0.9999, 0.5013)$$

with

$$RMSE_{drift} = 0.0801, \qquad RMSE_{diff} = 0.01052$$

and the corresponding optimal parameters

$$\alpha_{drift} = \alpha_{diff} = 10^{-4}.$$

The results of Lasso regression for both, the Drift term and the Diffusion terms are quite good. The Diffusion term is estimated very well and is very close to the true functional form $B^2(x) = 0.5x^2 + x + 0.5$. The Lasso regression results for the Drift term are slightly worse, but these results convincingly suggest that the dominant power in the Lasso regression for the Drift term should be $x^3$ ($a_{drift}^{Lasso}(3) = -0.866$). However, the coefficient for the $x^5$ is also significant ($a_{drift}^{Lasso}(5) = -0.26$). Perhaps these results can be improved by eliminating some terms with coefficients close to zero (e.g., enforcing that $a(0) = a(1) = 0$ in the definition of $f_l(\mathbf{a}, x)$ in (46) for the Lasso drift regression).

In addition, these results might be improved if we consider a larger domain for estimating the Drift coefficient. We would like to recall that the Drift and Diffusion terms are estimated on the interval $[-L, L]$ with $L = 0.5$ for the cubic model. Clearly, the effect of the term $x^5$ is quite negligible compared to the term $x^3$ on this interval. Therefore, increasing $L$ might provide more sensitivity for the Lasso regression and alleviate this problem.

### Ridge Regression.

In this section, we use Ridge regression to shrink the regression coefficients by imposing a penalty term of coefficients squared. The optimization problem becomes

$$\hat{\mathbf{a}} = \arg\min_{\mathbf{a}} \frac{1}{NB} \sum_{k=1}^{NB} \Big( f_l(\mathbf{a}, x_k) - \hat{F}(x_k) \Big)^2 + \alpha \sum_{i=0}^{l} |a_i|^2. \tag{51}$$

Similar to the Lasso regression we consider $(M, \Delta t) = (1000, 0.01)$ and $\Delta x = 0.0125$ for the cubic model. The results for the Ridge regression are presented below.

$$\hat{\mathbf{a}}_{drift} = (0, 0, 0, 0, -0.9036, -0.0105, -0.0191, -1.7 \times 10^{-4}),$$

$$\hat{\mathbf{a}}_{diff} = (0, 0, 0, 0, 0, 0.484, 0.9961, 0.5011)$$

with

$$RMSE_{drift} = 0.0796, \qquad RMSE_{diff} = 0.01173$$

and the corresponding optimal parameters

$$\alpha_{drift} = 0.0001, \qquad \alpha_{diff} = 0.01.$$

We can see that the Ridge regression works very well in this case and is superior to the Lasso regression. In particular, the ridge regression is able to correctly identify the functional form of the Drift and Diffusion estimators. In contrast with the Lasso regression, here we can clearly see that the highest power for the Drift term is $x^3$ with $a^{Ridge}_{drift}(3) = -0.9036$ which agrees well with $A(x) = -x^3$. For the Ridge regression coefficient of $x^5$ for the Drift is $a^{Ridge}_{drift}(5) = O(10^{-14})$. While the RMSE for the Drift term are comparable for the Lasso and Ridge regression, Ridge regression is clearly superior since it correctly identifies the functional form of the Drift term. The Diffusion term is also estimated very well here ($B^2(x) = 0.5x^2 + x + 0.5$).

### Lasso and Ridge Regression for $\Delta t = 0.001$.

Here we present the results for the Lasso and Ridge regression for the Drift and Diffusion estimators for the cubic model computed with $(M, \Delta t) = (1000, 0.001)$. As we saw in section 3.5.1 on the polynomial regression for the cubic model, the Drift estimator is computed considerably less accurately for $\Delta t = 0.001$ and thus, polynomial regression could not correctly identify coefficients of the Drift term (see Table 6 for $(M, \Delta t) = (1000, 0.001)$). Thus, we want to verify whether Lasso and Ridge regression techniques would mitigate this drawback of estimation for small $\Delta t$.

The results for $(M, \Delta t) = (1000, 0.001)$ are presented below.

$$\hat{\mathbf{a}}^{Lasso}_{drift} = (0, 0, -0.6247, 0.112, -0.3843, 0, -0.063, -2.7 \times 10^{-3})$$

$$\hat{\mathbf{a}}^{Lasso}_{diff} = (-0.016, -0.02, -2.3 \times 10^{-4}, -9.2 \times 10^{-3}, 2.5 \times 10^{-3}, 0.5015, 0.999, 0.4999)$$

with $RMSE^{Lasso}_{drift} = 0.24803$ and $RMSE^{Lasso}_{diff} = 0.0108$ (with $\alpha_{drift} = \alpha_{diff} = 10^{-4}$).

$$\hat{\mathbf{a}}^{Ridge}_{drift} = (-2.098, -0.131, -0.833, 0.124, -0.259, 0.016, -0.063, -0.004)$$

$$\hat{\mathbf{a}}_{diff}^{Ridge} = (0, 0, 0, 0, 0, 0.4988, 0.9991, 0.500)$$

with $RMSE_{drift}^{Ridge} = 0.2485$ and $RMSE_{diff}^{Ridge} = 0.01087$ (with $\alpha_{drift} = 10^{-4}$, $\alpha_{diff} = 10^{-2}$).

Both techniques are able to estimate the Diffusion term correctly. It is easy to see that the results of both, Lasso and Ridge regression, suggest that the diffusion term is very close to $B^2(x) = 0.5x^2 + x + 0.5$. However, both techniques fail for correctly estimating the Drift term $A(x) = -x^3$. Both, Lasso and Ridge regression, suggest that the highest power in the Drift term should higher than $x^3$. For the Lasso regression, the highest power is $x^5$ and for the Ridge regression the highest power is $x^7$. The corresponding regression coefficients for the Lasso and Ridge are $a_{drift}^{Lasso}(5) = -0.6247$ and $a_{drift}^{Ridge}(7) = -2.098$, respectively. The results of this and previous sections on Lasso and Ridge regression suggest that it is crucial to select correct (larger) $\Delta t$ for computing the Drift estimator $\hat{A}(x_k)$. In contrast to the Drift estimator, the Diffusion estimator is less sensitive to the selection of $\Delta t$.

# Bibliography

[1] Bernt Øksendal, Stochastic differential equations, 5th Edition, Springer, 2000.

[2] Hannes Risken, The Fokker-Planck equation: methods of solution and applications, 2nd Edition, Springer, 1996.

[3] Grigorios A. Pavliotis, Stochastic processes and applications: diffusion processes, the Fokker-Planck and Langevin equations, Springer, 2014.

[4] Philip Sura, Joseph Barsugli, A note on estimating drift and diffusion parameters from time-series, Physics Letters A, 304-311, 2002.

[5] Gerald B. Folland, Real analysis: modern techniques and their applications, 2nd Edition, John Wiley & Sons, 1999.

[6] Lacus, S. M, Simulation and inference for stochastic differential equations with R examples, Springer, 2008.

[7] C.W. Gardiner, Handbook of stochastic methods, 3rd Edition, Springer, 2004.

[8] S. Siegert, R. Friedrich, J. Peinke, Analysis of data sets of stochastic systems, Physics Letters A, 1998.

[9] R. Friedrich, S. Siegert, J. Peinke, St. Lück, M. Siefert, M. Lindeman, J. Raethjen, G. Deusch, G. Pfister, Extracting model equations from experimental data, Physics Letters A, 2000.

[10] R. Friedrich, S. Siegert, Ch. Renner, How to qualify deterministic and random influences on the statistics of the foreign market, Physics Letters A, 2000.

[11] J. Gradišek, S. Siegert, R. Friedrich, I. Grabec, Analysis of time series from stochastic processes, Physical Review E, 2000.

[12] Lawrence C. Evans, An introduction to stochastic differential equations , AMS, 2013.

[13] Jeffrey S. Rosenthal, A first look at rigorous probability theory, 2nd Edition, World Scientific Publishing Co. Pte. Ltd, 2006.

[14] Peter E. Kloeden, Eckhard Platen, Numerical solution of stochastic differential equations, Springer, 1992.

[15] Pagan, A. R. and Ullah, A. Non-parametric econometrics, Cambridge University Press, 1997.

[16] Lorenzo Boninsegna, Feliks Nuske, and Cecilia Clementi. Sparse learning of stochastic dynamics. J. Chem. Phys., 2018.

[17] F. Legoll and T. Lelievre. Effective dynamics using conditional expectations, Nonlinearity, 2010.

[18] W. Zhang and C. Schutte. Reliable approximation of long relaxation timescales in molecular dynamics, Entropy, 2017.

[19] M. Ratto, A. Pagano, P. C. Young, Non-parametric estimation of conditional moments for sensitivity analysis, Reliability Engineering & System Safety, 2009.

[20] Ashesh Chattopadhyay, Ebrahim Nabizadeh, and Pedram Hassanzadeh. Analog forecasting of extreme-causing weather patterns using deep learning. arXiv:1907.11617, 2019.

[21] S. Kravtsov, D. Kondrashov, and M. Ghil. Multilevel regression modeling of nonlinear processes: derivation and applications to climatic variability. J. Climate, 2005.

[22] P. Sura. Stochastic analysis of Southern and Pacific Ocean sea surface winds. J. Atmos. Sci., 2003.

[23] A. C. Lorenc, Analysis methods for numerical weather prediction, Quarterly Journal of the Royal Meteorological Society, 1986.

[24] J C Wyngaard, Atmospheric turbulence, Annual Review of Fluid Mechanics, 1992

[25] Tilmann Gneiting, Adrian E. Raftery, Weather forecasting with ensemble methods, Science, 2005.

[26] D. Crommelin and E. Vanden-Eijnden. Fitting timeseries by continous-time markov chains: A quadratic programming approach. J. Comp. Phys., 2006.

[27] David John Gagne II, Hannah M. Christensen, Aneesh C. Subramanian, and Adam H. Monahan. Machine learning for stochastic parameterization: Generative adversarial networks in the Lorenz 96 model. arXiv:1909.04711, 2019.

[28] F. Lu, K. K. Lin, and A. J. Chorin. Data-based stochastic model reduction for the Ku- ramotoSivashinsky equation. Physica D, 2017.

[29] K. Nimsaila and I. Timofeyev. Markov chain stochastic parametrizations of essential variables. SIAM Mult. Mod. Simul., 2010.

[30] Pantelis R. Vlachas, Jaideep Pathak, Brian R. Hunt, Themistoklis P. Sapsis, Michelle Girvan, Edward Ott, and Petros Koumoutsakos. Forecasting of spatio-temporal chaotic dynamics with recurrent neural networks: a comparative study of reservoir computing and backpropagation algorithms. arXiv:1910.05266, 2019.

[31] Feliks Nuske, Peter Koltai, Lorenzo Boninsegna, and Cecilia Clementi. Spectral properties of effective dynamics from conditional expectations. arxiv:1901:01557, 2019.

[32] B Mohammadi, O Pironneau, Analysis of the k-epsilon turbulence model, Editions MASSON, 1993.

[33] P Spalart, S Allmaras, A one-equation turbulence model for aerodynamic flows, 30th Aerospace Sciences Meeting and Exhibit, 1992.

[34] TT Warner, Numerical weather and climate prediction, Cambridge University Press, 2010

[35] M Mu, D Wansuo, W Jiacheng, The predictability problems in numerical weather and climate prediction, Advances in Atmospheric Sciences, 2002.

[36] AD Moura, S Hastenrath, Climate prediction for Brazil's Nordeste: performance of empirical and numerical modeling methods, Journal of Climate, 2004.

[37] EH Mller, R Scheichl, Massively parallel solvers for elliptic partial differential equations in numerical weather and climate prediction, Quarterly Journal of the Royal Meteorological Society, 2014.

[38] P Bauer, A Thorpe, G Brunet, The quiet revolution of numerical weather prediction, Nature, 2015.

[39] Niya Chen, Zheng Qian, Ian T. Nabney, Xiaofeng Meng, Wind power forecasts using Gaussian processes and numerical weather prediction, IEEE, 2013.

[40] Clifford F. Mass, David Ovens, Ken Westrick, and Brian A Colle, Does increasing horizontal resolution produce more skillful forecasts?: The results of two years of real-time numerical weather prediction over the pacific northwest, AMS, 2002.

[41] RJ Kuligowski, AP Barros, Localized precipitation forecasts from a numerical weather prediction model using artificial neural networks, Weather and Forecasting, 1998.

[42] Gary M. Carter, J. Paul Dallavalle, and Harry R. Glahn, Statistical forecasts based on the national meteorological center's numerical weather prediction system, AMS, 1989.

[43] Federico Cassola, Massimiliano Burlando, Wind speed and wind energy forecast through Kalman filtering of Numerical Weather Prediction model output, Applied Energy, 2012.

[44] B Brasnett, A global analysis of snow depth for numerical weather prediction, Journal of Applied Meteorology, 1999.

[45] JC Dallon, Numerical aspects of discrete and continuum hybrid models in cell biology, Applied Numerical Mathematics, Elsevier, 2000.

[46] James C. Schaff, Fei Gao, Ye Li, Igor L. Novak, and Boris M. Slepchenko, Numerical approach to spatial deterministic-stochastic models arising in cell biology, PLoS Comput Biol, 2016.

[47] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, Greta M Ljung, Time series analysis: forecasting and control, John Wiley & Sons, 2015.

[48] Francis X Diebold, Glenn D Rudebusch, Forecasting output with the composite leading index: A real-time analysis, Journal of the American Statistical Association, 1991.

[49] Hui Ding, Goce Trajcevski, Peter Scheuermann, Xiaoyue Wang, Eamonn Keogh, Querying and mining of time series data: experimental comparison of representations and distance measures, VLDB Endowment, 2008.

[50] Michael Clements, David Hendry, Forecasting economic time series, Cambridge University Press, 1998.

[51] Jan G De Gooijer, Rob J Hyndman, 25 years of time series forecasting, Elsevier, 2006.

[52] John A. Rice, Mathematical statistics and data analysis, Cengage Learning, 2007.

[53] Trevor Hastie, Robert Tibshirani, Jerome Friedman, The elements of statistical learning data mining, inference, and prediction, Springer, 2013.

[54] R. Lyman Ott , Micheal T. Longnecker, An Introduction to statistical methods and data analysis, Cengage Learning, 2015

[55] Allen Downey , Think Stats: Exploratory data analysis, O'Reilly Media, 2014.

[56] George Casella, Roger L Berger, Statistical inference, Cengage Learning, 2001.

[57] Robert Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society: Series B (Methodological), 1996.

[58] Donald W. Marquardt, Ronald D. Snee, Ridge regression in practice, The American Statistician, 1975.

[59] I. V. Basawa, A. K. Mallik, W. P. McCormick and R. L. Taylor, Bootstrapping explosive autoregressive processes, The Annals of Statistics, 1989.

[60] Yoon-JinLee, RyoOkui and MototsuguShintani, Asymptotic inference for dynamic panel estimators of infinite order autoregressive processes, Journal of Econometrics, 2018.

[61] Jerry Gibson, Entropy power, autoregressive models, and mutual information, Entropy,2018

[62] Christoph Bergmeir, Rob J.Hyndman and Bonsoo Koo, A note on the validity of cross-validation for evaluating autoregressive time series prediction, Computational Statistics & Data Analysis, 2018.

[63] Han Li, Kai Yang, Shishun Zhao and Dehui Wang, First-order random coefficients integer-valued threshold autoregressive processes, AStA Advances in Statistical Analysis, 2018.

[64] CHO, J. H., East asian financial contagion under DCC-Garch. International Journal of Banking and Finance, 2020.

[65] Jarociski, Marek, and Peter Karadi, Deconstructing monetary policy surprisesthe role of information shocks, American Economic Journal: Macroeconomics, 2020.

[66] TRCK, Stefan. Modelling and forecasting volatility in the gold market. International Journal of Banking and Finance, 2020.

[67] Cavit Pakel, Neil Shephard, Kevin Sheppard and Robert F. Engle, Fitting vast dimensional time-varying covariance models, Journal of Business & Economic Statistics,2020.

[68] Tim Bollerslev, Andrew J. Patton, Rogier Quaedvlieg, Multivariate leverage effects and realized semicovariance GARCH models, Journal of Econometrics, 2020.

[69] Steven L. Brunton, Joshua L. Proctor and J. Nathan Kutz, Sparse Identification of nonlinear dynamics with control (SINDYc), IFAC (International Federation of Automatic Control), 2016.

[70] Samuel H. Rudy,, Steven L. Brunton, Joshua L. Proctor and J. Nathan Kutz, Data-driven discovery of partial differential equations, Science Advances, 2017.

[71] Yohai Bar-Sinai, Stephan Hoyer, Jason Hickey, and Michael P. Brenner, Learning data-driven discretizations for partial differential equations, PNAS, 2019.

[72] P. Richard Hahn, Jared S. Murray, and Carlos M. Carvalho, Bayesian regression tree models for causal inference: regularization, confounding, and heterogeneous effects, International Society for Bayesian Analysis, 2020.

[73] Munish Kumar, , N. K. Tiwari and Subodh Ranjan, Prediction of oxygen mass transfer of plunging hollow jets using regression models, ISH Journal of Hydraulic Engineering, 2020.

[74] Ngoc Khue Tran, LAN property for an ergodic Ornstein Uhlenbeck process with Poisson jumps, Communications in Statistics, 2017.

[75] Nina Munkholt Jakobsen, Michael Srensen, Estimating functions for jump-diffusions, Stochastic Processes and their Applications, 2019.

[76] Gregor Pasemann and Wilhelm Stannat, Drift estimation for stochastic reaction-diffusion systems, Electron. J. Statist., 2020.

[77] Shimizu, Y., Some remarks on estimation of diffusion coefficient for jump-diffusions from finite samples, Bull. Inform. Cybernet, 2008.

[78] Shimizu, Y. and Yoshida, N, Estimation of parameters for diffusion processes with jumps from discrete observations. Statistical Inference for Stochastic Processes, 2006.