10-4-2023

Inference in Generalized Mean Reverting Processes

Yunhong Lyu
University of Windsor

Inference in Generalized Mean Reverting Processes

by

Yunhong Lyu

A Dissertation

Submitted to the Faculty of Graduate Studies

through the Department of Mathematics and Statistics

in Partial Fulfillment of the Requirements for

the Degree of Doctor of Philosophy at the

University of Windsor

Windsor, Ontario, Canada

2023

Inference in Generalized Mean Reverting Processes

by

Yunhong Lyu

APPROVED BY:

F. Viens, External Examiner
Rice University

A. Ngom
School of Computer Science

A. Hussein
Department of Mathematics and Statistics

M. Belalia
Department of Mathematics and Statistics

S. Nkurunziza, Advisor
Department of Mathematics and Statistics

September 15, 2023

Declaration of Co-Authorship/Previous Publications

I. Co-Authorship

I hereby declare that this thesis incorporates material that is result of joint research, as follows: Chapter 2, 3, and 4 of the thesis were co-authored with professor Sévérien Nkurunziza. In all cases, the primary contributions, simulation, data analysis, interpretation, and writing were performed by the author, and the contribution of co-authors was primarily through the provision of some theoretical results.

I am aware of the University of Windsor Senate Policy on Authorship and I certify that I have properly acknowledged the contribution of other researchers to my thesis, and have obtained written permission from each of the co-author(s) to include the above material(s) in my thesis.

I certify that, with the above qualification, this thesis, and the research to which it refers, is the product of my own work

II. Previous Publication

This thesis includes four original papers that have been previously published/submitted for publication in peer reviewed journals, as follows:

Thesis Chapter	Publication title/full citation	Publication status*
Chapter 2	Y. Lyu and S. Nkurunziza, (2023). Inference in Generalized Exponential O-U Processes. *Statistical Inference for Stochastic Processes.* **26**, 581–618	Published
	Y. Lyu and S. Nkurunziza, (2023). Inference in Generalized Exponential O-U Processes with Change-point *Statistical Inference for Stochastic Processes* $https://doi.org/10.1007/s11203-023-09293-z$	To appear
Chapter 3	Y. Lyu and S. Nkurunziza, (2023). Estimation and Testing in Generalized CIR Model *Annals of Applied Probability*	Under review

Thesis Chapter	Publication title/full citation	Publication status*
Chapter 4	Y. Lyu and S. Nkurunziza, (2023). Inference methods in time-varying linear diffusion processes *Bernoulli*	Under review

I certify that I have obtained a written permission from the copyright owner(s) to include the above published material(s) in my thesis. I certify that the above material describes work completed during my registration as a graduate student at the University of Windsor.

III. General

I declare that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. Furthermore, to the extent that I have included copyrighted material that surpasses the bounds of fair dealing within the meaning of the Canada Copyright Act, I certify that I have obtained a written permission from the copyright owner(s) to include such material(s) in my thesis. I declare that this is a true copy of my thesis, including any final revisions, as approved by my thesis committee and the Graduate Studies office, and that this thesis has not been submitted for a higher degree to any other University or Institution.

Abstract

This dissertation proposes three types of processes that are suitable for modeling positive datasets with periodic behavior and mean-reverting level phenomenon.

A class of generalized exponential Ornstein–Uhlenbeck process (GEOU) is considered in Chapter 2. This chapter's key characteristics include the following: first, the classical exponential Ornstein–Uhlenbeck process is generalized to the case where the drift coefficient is driven by a period function of time; second, as opposed to the results in recent literature, the dimension of the drift parameter is considered unknown. This chapter serves to weaken some assumptions, in recent literature, underlying the asymptotic optimality of some estimators of the drift parameter. Three types of estimators are proposed: unrestricted maximum likelihood estimator (UMLE), restricted maximum likelihood estimator (RMLE) and shrinkage estimators (SEs). Asymptotic distributional risk (ADR) of the proposed estimators is also derived, as well as their relative efficiency. Further, it is proven that the proposed methods improve the goodness-of-fit. Finally, this chapter outlines an analysis of a financial market data set and presents the simulation results, which corroborate the theoretical findings.

Chapter 3 proposes a generalized Cox–Ingersoll–Ross (GCIR) process that is suitable for modeling some periodic financial data. An inference problem, about the drift parameters of the introduced GCIR process is also considered when the target parameters may satisfy some restrictions. Like in the case of GEOU process, three kinds of estimators are derived: UMLE, RMLE, and SEs. Their joint asymptotic normality is studied. Based on the established asymptotic result, a test is constructed for testing the restriction. The asymptotic power of the proposed test is also derived from this, and it is proven that the proposed test is consistent. This chapter also outlines the ADR of the proposed estimators and their relative efficiency. Finally, simulation results are

presented that corroborate the study's theoretical findings.

In Chapter 4, a generalized Chan, Karolyi, Longstaff and Sanders (GCKLS) process is proposed for modeling some financial data that are cyclical in nature. The ergodicity of the solution to the GCKLS model is proven by using the transition probability; the normality and strong consistency of the UMLE are proven by using the ergodicity. Similarly, UMLE, RMLE, and SEs are derived. A test is performed to assess the restriction. The asymptotic power of the proposed test is consistent. Further, the relative efficiency of the proposed estimators is compared, and simulation results are presented that agree with our theoretical findings.

Dedication

Dedicated to my family

Acknowledgements

First, I am deeply grateful to my advisor, Dr. Sévérien Nkurunziza, for his unwavering support and guidance throughout my doctoral studies. The years I spent studying with Dr. Nkurunziza were very precious to me. Academically, Dr. Nkurunziza is very knowledgeable and far-reaching, and he is able to see the smallest details and understand them in depth, which always inspired me whenever I talked with him. Dr. Nkurunziza was gentle and humble. His strict academic spirit and optimistic and open-minded philosophy have always deeply influenced me. At the beginning of my studies, my lack of understanding of the topic and the slow progress of my research caused me a deep sense of frustration, which was relieved by the patient and systematical guidance of my supervisor. I would like to thank my supervisor for correcting my papers word by word. During the writing of my doctoral dissertation, you not only gave me professional advice on the overall framework and specific contents of my dissertation, but also gave me careful guidance on the format and details of my argumentation, which made my dissertation even better. Once again, I would like to thank my supervisor for his help and training, and I could not have completed this dissertation without his mentorship.

Further, I want to express my gratitude to my committee members. To begin, I would like to express my sincere appreciation to Dr. George Lan from Odette School of Business for serving as the Chair of my defense. I would also like to thank the external examiner, Dr. Frederi G. Viens from the Department of Statistics, Rice University and my advisory committee members Dr. Abdulkadir Hussein, Dr. Mohamed Belalia, and Dr. Alioune Ngom. I appreciate the time you have spent reading my dissertation and I am grateful for your valuable comments and suggestions. I would like to express my sincere gratitude and best wishes to all the staff in department of Mathematics and Statistics. Thank you for your help to me directly and indirectly during my Ph.D. study.

Last but not least, I would like to thank my family members for their silent care and support to keep me on the path of my studies. Thanks to my parents, my parents-in-law, and my husband, without their dedication and support, I would not have been able to complete my research. Thanks also to my lovely sons. They are my moral support, my source of optimism and strength to keep on going!

Thank you to all those who played a role in making this dissertation possible.

Table of Contents

List of Tables

List of Figures

Page

Chapter 1

General Introduction

Ordinary differential equations are widely used in solving problems in the fields of engineering, physics, biology, and economics. However, the real world is inevitably affected by certain stochastic factors. Therefore, the analysis of practical problems needs to be extended from a deterministic to a stochastic point of view. Thus, stochastic differential equations (SDE) have come to our attention and gradually gained the interest of a large number of researchers. At the beginning of the 20th century, Einstein [1905] established the mathematical theory of Brownian motion and molecular diffusion, which has since been applied in various fields such as chemical kinetics, population genetics, social sciences and engineering. However, the study of SDEs did not go smoothly. The Itô equation (Itô [1951]) is an important method for studying the SDEs whose solutions are Markov processes, and therefore it is of great significance for the study of stochastic process theory and control theory. As stated in Ibe [2013], diffusion processes are continuous-time, continuous-state processes whose sample paths are everywhere continuous but nowhere differentiable. Nowadays, diffusion processes are mainly used to model physical, biological, engineering, economic, and social phenomena because diffusion is one of the fundamental mechanisms for the transport of materials in physical, chemical, and biological systems.

In practical applications, the parameters of the diffusion processes are totally or partially unknown due to the interference of random factors. Therefore, the estimation of unknown parameters has become a critical problem to be solved in order to better understand the relevant asset dynamics. In the past decades, scholars have studied the parameter estimation problem based on economic models represented by continuous time diffusion processes, and have achieved some remarkable results. To give some references, see Koroliuk et al. [2020], Kubilius et al. [2017], Favetto [2014].

This dissertation is focusing on economic probability modeling. In an era characterized by economic uncertainty and fluctuation, the research conducted in this dissertation holds a strong significance, as it directly addresses the pressing societal need for robust tools to model and predict economic probabilities. The economic models studied in this dissertation are widely used in quantitative finance for modeling asset prices, interest rates, derivatives pricing, stock prices, and macroeconomic dynamics. To give some examples, we quote Behme and Sideris [2022], Liu et al. [2006], Nowak and Romaniuk [2014], Dassios et al. [2019], Ben Nowman and Sorwar [2003], Khor et al. [2012] and references therein.

This dissertation considers observing a stochastic process $\{X(t), t \geqslant 0\}$ which is a solution of the SDE

$$dX(t) = S(\theta, t, X(t))dt + \sigma (X(t))^{\delta} dB_t, \tag{1.0.1}$$

where $\{B_t, t \geqslant 0\}$ is a Brownian motion and $S(\theta, t, X(t))$ is a function of t, θ and $X(t)$. σ is the known parameter associated with volatility. The parameter δ determines the sensitivity of the variance to the level of the process $X(t)$. If $S(\theta, t, X(t)) = \alpha(\mu - X(t)), \delta = 0$, the process is the well known *Uhlenbeck-Ornstein* (O-U) process. If $S(\theta, t, X(t)) = \alpha(\mu - \ln X(t))X(t)$, α, μ are the parameters to be estimated and $\delta = 1$, the process is the well-known exponential *Uhlenbeck-Ornstein* process; while if $S(\theta, t, X(t)) = \beta - \alpha X(t)$ with α, β to be estimated and $\delta = 1/2$, the process becomes the classical *Cox–Ingersoll–Ross* (CIR) model; Further, if $S(\theta, t, X(t)) = \beta - \alpha X(t)$, and the constant $\delta > 0$, the process is considered as *Chan, Karolyi, Longstaff and Sanders* (CKLS) model.

Parameter estimation is a critical aspect of stochastic differential equations (SDEs) that holds significant implications across various fields of science and engineering. SDEs are invaluable tools for modeling systems influenced by random fluctuations, from financial markets to ecological ecosystems. Accurate estimation of the parameters governing these equations is fundamental for gaining insights, making predictions, and informed decision-making.

In the realm of finance, parameter estimation in SDEs is pivotal for risk management, portfolio optimization, and derivative pricing. Models like the Black-Scholes equation, driven by SDEs, underpin options pricing and risk assessment in financial markets. Accurate parameter estimation ensures that investors and financial institutions can better understand and mitigate risks, ultimately contributing to financial stability and sound decision-making.

To underscore the importance of parameter estimation in SDEs, it is essential to reference foundational texts like Oksendal [2013], Nielsen et al. [2000]. These references provide comprehensive insights into the theory and applications of SDEs, emphasizing the critical role of parameter estimation in harnessing the predictive power of these equations.

There are many publications regarding the parameters estimation of the exponential O-U process, CIR process and CKLS process. To give some references, see Kellerhals [2013], Vega [2018], Feng and Xie [2012], Alaya and Kebaier [2012], Alaya and Kebaier [2013], Chen and Scott [1993], Wei [2020] and some references therein.

In the context of these models, inference problem about the drift parameters has been addressed to some extent. The common factor about these processes consists in the fact that they are suitable for the datasets which exhibit a constant mean-reverting level. However, the assumption of a constant mean level is not adequate due to seasonality patterns or a long-term trend of the process. Dehling et al. [2010] extended the O-U process to the case where the reversion term is a deterministic periodic function of time t. Dehling et al. [2014] considered the change-point detecting problem under the situation of periodic mean reversion process. By the explicit representation of the generalized likelihood ratio test statistic, this paper determined the asymptotic distribution of the test statistic under the null hypothesis. Nkurunziza and Zhang [2018] took the hypothesis testing problem a step further. The drift parameter was supposed to satisfy some linear restrictions. This introduces a more intricate dimension to the problem, where identifying the presence of a change-point emerges as a special case. Nkurunziza and Fu [2019] generalized the O-U process to the case with multiple change-points. Nkurunziza and Shen [2019], Nkurunziza [2021] generalized the O-U process to the multivariate case. By combining the results in Nkurunziza [2015], Nkurunziza and Ahmed [2010], other authors such as Nkurunziza and Zhang [2018], Nkurunziza and Fu [2019], Nkurunziza and Shen [2019], Nkurunziza [2021] considered shrinkage estimators. Thus, strongly motivated by these cited papers, a more general cases is considered with a deterministic and periodic drift term in all the three different types of SDEs described above. Particularly, the inference problem regarding the drift parameters catches the interest.

This thesis is organized as follows: In Chapter 2, the generalized exponential O-U (GEOU) model is proposed. The estimation and hypothesis testing problem are considered under the case where the drift term is a periodic function and satisfies some linear restrictions. The change-point problem is also studied. After deriving the explicit solution, a stationary and ergodic stochastic process is constructed. The distance between the constructed process and the solution converges to 0 almost surely and in mean as time tends to infinity. The convergence of this distance implies the asymptotic normality of the estimators and the consistency of the hypothesis test. In this chapter, unlike the settings in Nkurunziza and Zhang [2018], Nkurunziza and Fu [2019], Nkurunziza and Shen [2019], Nkurunziza [2021], the dimension p of the mean reversion is supposed to be unknown. Through the utilization of hypothesis testing, both the dimension p and the presence of a change-point t^* are identified, which has led to an enhancement in the accuracy of the predictions. The comparison of the predicting accuracy is also given in this chapter. Based on the estimation of dimension p and change-point t^*, estimates of the drift parameters are derived. The asymptotic normality of the proposed estimators as well as the consistency of the hypothesis test are given.

In Chapter 3, we propose the generalized *Cox–Ingersoll–Ross* model (GCIR). The main novelty of this chapter is that a stationary and ergodic process is constructed despite the absence of an explicit solution of the GCIR model. To prove the constructed auxiliary process is stationary and ergodic, we use the extension of Dubins-Schwarz theorem. Like in Chapter 2, it is also proven that the distance between the auxiliary process and the implicit solution of GCIR process converges to 0 almost surely and in mean. The asymptotic normality of the estimators are also derived.

In Chapter 4, the generalized Chan, Karolyi, Longstaff and Sanders (GCKLS) process is proposed. The proposed process is used for modeling some financial data that are cyclical in nature. It should be noted that the GCKLS process is a generalization of the GCIR process proposed in Chapter 3. Under the context of GCKLS process, the sensitive parameter belongs to $(1/2, 1)$ rather than a constant $1/2$ as proposed in GCIR process. In GCKLS model, the long-run drift term is a periodic deterministic function rather than a constant. By using the transition probability, the ergodicity of the solution to the generalized CKLS model is obtained. Based on the ergodicity, strong consistency and asymptotic normality of the UMLE are proven. It is important to observe that, in the special case where the periodic base function is analytic on the top of satisfying other conditions of Assumption 3.2, the ergodicity of the GCIR model discussed in Chapter 3 can be derived using the transition probability approach used in this broader class - GCKLS model, with the sensitive parameter of 1/2. However, the current investigations do not allow us to use the construction of auxiliary processes to solve the problem in Chapter 4. On the other hand, without assuming that the base periodic function is analytic with further restrictions as in Assumption 4.2, current investigations do not allow us to use the method in Chapter 4 in order to solve the inference problem in GCIR described in Chapter 3.

For all the three types of SDEs, first, three types of estimators are derived: maximum likelihood estimator with no prior information (UMLE), the maximum likelihood estimator under some given restriction (RMLE), and some shrinkage estimators (SEs). As described in Nkurunziza [2015], shrinkage estimators (SEs) combine in an optimal way the UMLE and the RMLE. As frequently noticed in constrained inference, if the restriction is not correct, the UMLE performs better than the RMLE while if the restriction holds, the RMLE dominates the UMLE. However, as in Nkurunziza [2015], more often than not, it is not possible to be totally sure about the validity of the restriction. Thus, it is important to derive a statistical method which is robust with respect to the restriction. The SEs have the advantage of preserving a very good performance regardless of the validity of the restriction. Nevertheless, since the dimensions of the UMLE and the RMLE in GEOU process are random, the derivation of shrinkage estimators as well as their relative efficiency do not follow from the results in classical literature. Based on their joint asymptotic normality, a test for assessing the restrictions is constructed. Furthermore, the study establishes the consistency of the test. It includes a thorough analysis comparing the effectiveness of various types of estimators. Ultimately, the simulation results not only validate the theoretical discoveries but also underscore the appeal of the method under examination

Chapter 5 summarized this dissertation and gave potential area of future research.

CHAPTER 2

Inference in GEOU Process

2.1 Introduction

The basic commodities essentials of humanity, such as crude oil, natural gas, gold, silver, corn, wheat, etc. play an important role in both keeping sustainability and improving civilization. In worldwide financial markets, changes in commodities' prices can have a huge impact on human life. As mentioned in [Schwartz, 1997, Casassus and Collin-Dufresne, 2005], one of the most recognized highlights of commodities' price is that the price possesses a mean-reverting behavior. Many statistical models have been established in light of this property of commodities' prices. For instance, the most basic and simplified stochastic process that describes the characteristic of the process to drift toward a long-term value is known as the Ornstein–Uhlenbeck process. Later, the Ornstein–Uhlenbeck process is denoted as O-U process for short. The classical O-U process $\{X(t), t \geqslant 0\}$ is the solution of the stochastic differential equation (SDE) $dX(t) = (\alpha X(t) + \beta)dt + \sigma dB_t$, where α, β, σ are constants, $\{B_t, t \geqslant 0\}$ is a standard Brownian motion. This SDE is one of a few cases that admits an explicit solution. However, in practice, the assumption of a constant mean reversion is rarely met. To address this issue, Aalen and Gjessing [2004] extended the constant mean reversion level to a time-varying mean reversion function. Further, Dehling et al. [2010] proposed a generalized OU process with a periodic mean reversion, given by $dX(t) = (L(t) - \alpha X(t))dt + \sigma dB_t, \;\; t \geqslant 0$, where $L(t)$ is a time-varying periodic mean reversion level and α, σ are positive constants. Recently, Nkurunziza and Zhang [2018] studied the inference problem about the drift parameter in the generalized mean-reverting process with a change-point under uncertain linear restriction. To give other related references, see Chen et al. [2018] who developed some estimation methods for the change-point.

However, one of the main limitations of the cited works consists in the fact that, in the models considered, the process can take a negative value while in fact, the observations such as spot prices cannot have negative values. To address this limitation, Dixit and Pindyck [1994] developed the so-called geometric O-U process, and Schwartz [1997] proposed a stochastic process known as exponential O–U process or *Schwartz process*. Namely, an exponential O–U process is a solution of the SDE

$$dX(t) = \alpha(\mu - \ln X(t))X(t)dt + \sigma X(t)dB_t. \tag{2.1.1}$$

In particular, this process can be used for modeling the spot price of the commodity. In this case, the magnitude of the speed of adjustment $\alpha > 0$ measures the degree of mean reversion to the long-run mean log price. One reference is Masoliver and Perelló [2006] who studied the exponential O–U stochastic volatility model and observed that the model shows a multiscale behavior in the volatility autocorrelation. Another related reference is Perelló et al. [2008] who analyzed the pricing issue for a European call choice when the volatility of the underlying asset follows the exponential O–U process. Just recently, Vega [2018] presented a methodological procedure to estimate the parameters of the exponential O–U process and gave a comparison between the MLE and least squares estimator.

As for the classical O-U process, the exponential O-U process is suitable for modeling the datasets for which the mean reverting level is a constant. For instance, such processes are not inadequate in modeling the data with seasonality patterns. In this chapter, a more general process that is suitable for modeling nonnegative datasets with time-varying periodic mean-reverting behavior and possible drastic changes is considered. Thereafter, such a process will be designated as a ”generalized exponential O-U (GEOU) process”. Further, the inference problems about the drift parameter vector of the GEOU was studied, in the context where the dimension of the parameter is unknown and under uncertain prior information about the target parameter. More precisely, the prior information in the form of linear restriction binding the components of the drift parameter is considered. So far, very little attention has been paid to the estimation of the drift parameter's dimensions. Another novelty in this chapter consists of the fact that a test is proposed to assess the restrictions, as well as the parameter dimension. It is proven that the proposed methods improve the goodness-of-fit. To the best of the author's knowledge, there does not exist a similar research work in the context of GEOU. In summary, the main contributions of this chapter are as follows:

1. A process with mean-reverting level which is a periodic function of time t was introduced and this generalizes the so-called exponential O-U process.

2. By considering the case where the dimension of the drift parameter is unknown, the inference methods in Dehling et al. [2010], Dehling et al. [2014], Chen et al. [2018] and Nkurunziza and Zhang [2018] among others were improved. Incorporating an unknown dimension renders the process more aligned with reality.

3. The estimation and a testing problem about the drift parameter were studied. In particular, the UMLE, RMLE, and the SEs are proposed, as well as an asymptotic

test for testing the restriction. The optimality of the proposed test is studied, as well as the relative efficiencies of the proposed estimators. To this end, the difficulties are overcome due to the fact that the dimensions of these estimators are random.

4. It is proven that the proposed method improves the goodness-of-fit.

The rest of this chapter is structured as follows. Section 2.2 gives the statistical model and some preliminary results. In Section 2.3, inference problems in the case where the dimensions of the drift parameter are known are studied. Section 2.4 extends the inference methods to the case where the dimensions of the drift parameter are unknown. In Section 2.5, the relative efficiency of the proposed estimators are compared. Section 2.6 is the empirical study and numerical results. In this section, the proposed method is applied to the financial market historical dataset. Finally, Section 2.7 is the conclusion and, for the convenience of the reader, some theoretical results and proofs are given in Appendix A.

2.2 Statistical model and preliminary results

This section presents the statistical model and some preliminary results, as well as some useful notations. About the notations, let $(\Omega, \mathscr{F}, \mathrm{P})$ be a probability space where $\mathscr{F}$ is a $\sigma-$field on the sample space Ω, and P is a probability measure. Let $\{\mathscr{F}_t, t \geqslant 0\}$ denote the natural filtration associated to a standard Brownian motion $\{B_t\}_{t\geqslant 0}$. Further, let L^m denote the space of measurable $m-$integrable function, for some $m \geqslant 1$. Let $\xrightarrow[T\longrightarrow+\infty]{\mathrm{D}}, \xrightarrow[T\rightarrow+\infty]{L^m}, \xrightarrow[T\rightarrow+\infty]{a.s.}, \xrightarrow[T\rightarrow+\infty]{\mathrm{P}}$ be the convergence in distribution, in L^m-space, almost surely, and in probability, respectively, as T tends to infinity. Also, let $O_{\mathrm{P}}(a(T))$ stand for a random quantity such that $O_{\mathrm{P}}(a(T))a^{-1}(T)$ is bounded in probability and $o_{\mathrm{P}}(a(T))$

stand for a random quantity such that $o_{\mathrm{P}}(a(T))a^{-1}(T)$ converges in probability to 0 , as T tends to infinity. Let $\mathbb{R}_+ = [0, +\infty)$. Further, a stochastic process $\{X(t),\ t \geqslant 0\}$ is said to be L^m bounded if there exists $K > 0$ such that $\mathbb{E}(|X(t)|^m) < K$, for all $t \geqslant 0$. Thereafter, let I_n be the $n-$dimensional identity matrix and let $\mathbb{I}_A$ denote the indicator function of the event A, $\top$ denote the transpose of a matrix, $\mathbb{N}_+ = \{1, 2, 3, \cdots\}$ and $\|\cdot\|$ represent Frobenius norm of a matrix.

2.2.1 Statistical model

Inspired by the cited works, the statistical model under consideration is a generalization of model (2.1.1). The generalized exponential O-U process $\{X(t), t \geqslant 0\}$ is a stochastic process which satisfies the following stochastic differential equation

$$dX(t) = S(\theta, t, X(t))dt + \sigma X(t)dB_t, \quad X(0) = X_0 \tag{2.2.1}$$

where $S(\theta, t, X(t)) = (L(t) - \alpha \ln X(t))X(t)$, with $L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t)$, where, for each $i = 1, 2, \ldots, p$, the function $\varphi_i(t)$ is a real-valued function of t. Thereafter, the process in (2.2.1) will be referred to as the GEOU process. Let $\varphi(t) = \left(\varphi_1(t), \varphi_2(t), \varphi_3(t), ..., \varphi_p(t)\right)$, $t \geqslant 0$ and let $\theta = (\mu_1, \mu_2, \ldots, \mu_p, \alpha)^\top \in \Theta \subset \mathbb{R}^{p+1}$. In this chapter, θ is the target parameter and, as opposed to similar works in literature, p is an unknown nuisance parameter. Note that, in the continuous time observations, the diffusion parameter σ^2 can be consistently estimated by the discretized version of quadratic variation of the process $\{\ln X(t), t \geqslant 0\}$. Because of that and without loss of generality, σ^2 is supposed to be known. To give some closely related references in literature, see Dehling et al. [2014], Nkurunziza and Zhang [2018], and references therein. Nevertheless, as opposed to the quoted works, here the dimensions of the target parameter θ is unknown and the case where θ may satisfy some uncertain prior information is considered. In practice, the prior information may come from some the previous statistical investigations or from the fields' experts. In particular, θ is suspected to satisfy the following restriction:

$$H_0 : M\theta = r \tag{2.2.2}$$

where, for a fixed p, M is a known $q \times (p+1)$-full rank matrix with $q < p+1$, r is a known q column vector. In terms of estimation method, a statistical procedure which preserved a good performance regardless of the validity of the restriction in (2.2.2) is considered. Nevertheless, in order to validate the restriction in (2.2.2), a test is derived for the hypothesis testing problem

$$H_0 : M\theta = r \quad \text{versus} \quad H_1 : M\theta \neq r. \tag{2.2.3}$$

It should be noted that for a suitable choice of M, the null hypothesis in (2.2.3) corresponds to some interesting statistical problems. For instance, let $r = 0$, and $M = (I_p, -I_p)$, the restriction in (2.2.2) corresponds to the case where there are no change point. Thus, the testing problem in (2.2.3) includes as a special case testing the absence of change point. However, for the sake of clarity, the change-point case and the no change-point case are presented separately. The optimality of the proposed method relies on the following assumptions.

Assumption 2.1. The parameter $\alpha > 0$.

Assumption 2.2. For any $T > 0$, the base function $\varphi(t)$ is Riemann-integrable on [0,T] and possess

I Periodicity: $\varphi(t+\upsilon) = \varphi(t)$, for some period υ and all $t \in [0, T]$.

II Orthogonality in $L^2([0,\upsilon], \frac{1}{\upsilon}d\lambda)$: $\int_0^{\upsilon} \varphi(t)\varphi^{\top}(t)dt = \upsilon I_p$.

Remark 2.2.1. Since the base function $\varphi(t)$ is Riemann-integrable on [0,T] and υ-periodic, this implies that $\varphi(t)$ is bounded on $\mathbb{R}_+ = [0, +\infty)$, $\|\varphi(t)\| \leqslant K_{\varphi}$ for some positive constant K_{φ}.

Assumption 2.3. The distribution of the initial value, X_0, of the SDE in (2.2.1) does not depend on the drift parameter θ. Further, X_0 is positive a.s. and independent of $\{B_t : t \geqslant 0\}$ and $\mathbb{E}(|X_0|^m) < \infty$, $\mathbb{E}(|\ln X_0|^m) < \infty$, for some $m \geqslant 2$.

In this chapter, as in Dehling et al. [2014], Nkurunziza and Zhang [2018] and references therein, without loss of generality, the period υ is assumed to be known and equals to 1. In the case where p is known, the base function $\varphi(t)$ is also supposed to be known as in the quoted paper. Note that, for the case where p is unknown, the number of elements in base function $\varphi(t)$ need to be estimated. To simplify the presentation of this chapter, the case where p is known is treated separately from the case where p is unknown parameter.

2.2.2 Preliminary results

This subsection presents some preliminary results about the trajectory of the SDE in (2.2.1). It is proven that the SDE (2.2.1) possesses a strong and unique solution which is L^m-bounded. This result is essential in deriving the likelihood function, the unrestricted maximum likelihood estimator (UMLE) and the restricted maximum likelihood estimator (RMLE). The following proposition gives the representation of the solution to the SDE (2.2.1).

Proposition 2.2.1. *Suppose that Assumption 2.1-2.3 hold. Then, the solution of SDE (2.2.1) is given*

$$X(t) = \exp\left\{e^{-\alpha t}\ln X_0 + r(t) + \tau(t)\right\},$$

where $r(t) = e^{-\alpha t}\int_0^t e^{\alpha s}(L(s) - \frac{1}{2}\sigma^2)ds$ and $\tau(t) = \sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s$.

Proof. Let $V(t, X(t)) = e^{\alpha t}\ln X(t)$. By Itô's Lemma,

$$dV(t, X(t)) = \frac{\partial V}{\partial t}(t, X(t))dt + \frac{\partial V}{\partial X(t)}(t, X(t))dX(t) + \frac{1}{2}\frac{\partial^2 V}{\partial X^2(t)}(t, X(t))d\langle X(t), X(t)\rangle,$$

where $\langle \cdot, \cdot \rangle$ represent the variation. Since $dX(t) = (L(t) - \alpha \ln X(t))\, X(t)dt + \sigma X(t)dB_t$, then

$$\begin{aligned} dV(t, X(t)) =& \alpha e^{\alpha t} \ln X(t)dt + e^{\alpha t} \frac{1}{X(t)}((L(t) - \alpha \ln X(t))X(t)dt + \sigma X(t)dB_t) \\ & - \frac{1}{2} e^{\alpha t} \frac{1}{X^2(t)} \sigma^2 X^2(t)dt \\ =& e^{\alpha t} \left(L(t) - \frac{1}{2}\sigma^2 \right) dt + \sigma e^{\alpha t} dB_t, \end{aligned}$$

which implies that $d(e^{\alpha t} \ln X(t)) = e^{\alpha t} \left(L(t) - \frac{1}{2}\sigma^2 \right) dt + \sigma e^{\alpha t} dB_t$. Integrating both sides from 0 to t gives

$$e^{\alpha t} \ln X(t) = \ln X_0 + \int_0^t e^{\alpha s} \left(L(t) - \frac{1}{2}\sigma^2 \right) ds + \int_0^t \sigma e^{\alpha s} dB_s.$$

Then

$$\ln X(t) = e^{-\alpha t} \ln X_0 + e^{-\alpha t} \int_0^t e^{\alpha s} \left(L(t) - \frac{1}{2}\sigma^2 \right) ds + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s \qquad (2.2.4)$$

which implies that $X(t) = \exp\left\{ e^{-\alpha t} \ln X_0 + e^{-\alpha t} \int_0^t e^{\alpha s} \left(L(t) - \frac{1}{2}\sigma^2 \right) ds + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB_s \right\}$. This completes the proof. □

Proposition 2.2.1 shows that the solution of SDE (2.2.1) is positive with probability 1. This shows that the logarithm of $X(t)$ is well defined. Below, it is proven that the processes $\{\ln X(t), t \geqslant 0\}$ and $\{X(t), t \geqslant 0\}$ are L^m bounded. Such a result plays important role in deriving a sufficient condition for the likelihood function of the SDE (2.2.1).

Proposition 2.2.2. *If Assumption 2.1-2.3 hold, then,*

(1) $\sup_{t \geqslant 0} \mathbb{E}[|\ln X(t)|^m] < \infty$, (2) $\sup_{t \geqslant 0} \mathbb{E}[X(t)] < \infty$, *(3)* $\sup_{t \geqslant 0} \mathbb{E}[X^m(t)] < \infty$,

(4) $\sup_{t \geqslant 0} \mathrm{E} \left(|S(\theta, t, X(t))|^m \big/ |\sigma X(t)|^m \right) < \infty$,

(5) $\mathrm{P} \left(\int_0^T |S(\theta, t, X(t))|^m \big/ |\sigma X(t)|^m \, dt < \infty \right) = 1$, *for all* $0 \leqslant T < \infty$.

The proof is given in Appendix A.2. According to Theorem 7.6 [Liptser and Shiryaev, 2001, pp.261] , to get the likelihood function of a diffusion process,

$\mathbb{P}\left(\int_0^T S^2(\theta, t, X(t)) \Big/ (\sigma^2 X^2(t))\, dt < \infty\right) = 1$, for all $0 \leqslant T < \infty$ should be guaranteed. In Part **(5)** of the above proposition, the sufficient condition in a more general form is presented.

2.3 Inference in the case where p is known

This section studies inference problems in the case where the dimension of the parameters p is known. In particular, UMLE and RMLE are derived. Further, the joint asymptotic normality of these estimators is also derived, as well as an asymptotic test for the testing problem in (2.2.3).

2.3.1 The case of absence of change-point

2.3.1.1 Parameter estimation

This subsection derives the maximum likelihood estimator of θ. To this end, let $C_{[0,T]}$ be the space of continuous real-valued functions on $[0, T]$, let $\mathcal{B}_{[0,T]}$ be the associated Borel σ algebra and let P_X denote the probability measure induced by the observable realizations $X^T = \{X(t), t \geqslant 0\}$ on the measurable space $(C_{[0,T]}, \mathcal{B}_{[0,T]})$. Further, let P_B be the probability measure generated by the Brownian motion on $(C_{[0,T]}, \mathcal{B}_{[0,T]})$. Then, as in Dehling et al. [2014], dP_X/dP_B is the Radon-Nikodym derivative of the observations generated by the SDE (2.2.1). Further, the likelihood function of observations X^T is given by

$$\mathcal{L}(\theta, X^T) := \frac{dP_X}{dP_B}(X^T) = \exp\left\{\int_0^T \frac{S(\theta, t, X(t))}{\sigma^2 X^2(t)} dX(t) - \frac{1}{2}\int_0^T \frac{S^2(\theta, t, X(t))}{\sigma^2 X^2(t)} dt\right\}. \tag{2.3.1}$$

Thus, the maximum likelihood estimator (UMLE) $\hat{\theta}_T$ can be derived by taking the maximum value of the functional $\theta \longmapsto \mathcal{L}(\theta, X^T)$, i.e. $\hat{\theta}_T := \arg\max_{\theta} \mathcal{L}(\theta, X^T)$. To simplify some notations, let

$$Q_{[0,T]} = \begin{bmatrix} \int_0^T \varphi^\top(t)\varphi(t)dt & -\int_0^T \varphi^\top(t)\ln X(t)dt \\ -\int_0^T \varphi(t)\ln X(t)dt & \int_0^T (\ln X(t))^2 dt \end{bmatrix}, \tag{2.3.2}$$

$$U_{[0,T]} = \int_0^T \frac{(\varphi(t), -\ln X(t))^\top}{X(t)} dX(t), \qquad W_{[0,T]} = \int_0^T (\varphi(t), -\ln X(t))^\top dB_t. \tag{2.3.3}$$

Further, let $\tilde{\theta}_T$ be the RMLE. In deriving the UMLE and RMLE, the matrix $Q_{[0,T]}$ needs to be invertible. The following proposition gives a sufficient condition under which the matrix $Q_{[0,T]}$ is positive definite.

Proposition 2.3.1. *If Assumption 2.1-2.3 hold, then, if $T \geqslant 1$, $Q_{[0,T]}$ is a positive definite matrix.*

The proof of this proposition is provided in Appendix A.3. Since the optimality of the proposed method relies on the asymptotic properties, in the sequel, the condition $T \geqslant 1$ is always supposed to be true. Proposition 2.3.1 implies the following proposition which gives the UMLE and the RMLE. let $G_{[0,T]} = Q_{[0,T]}^{-1} M^\top (M Q_{[0,T]}^{-1} M^\top)^{-1}$.

Proposition 2.3.2. *Suppose that Assumptions 1-3 hold. Then,*

$$\hat{\theta}_T = \sigma Q_{[0,T]}^{-1} U_{[0,T]}, \quad \textit{and} \quad \tilde{\theta}_T = \hat{\theta}_T - G_{[0,T]}(M\hat{\theta}_T - r).$$

Proof. From Proposition 2.2.1 Part **(5)** and Theorem 7.6 in Liptser and Shiryaev [2001], the likelihood function of the SDE (2.1) is given by

$$\mathcal{L}(\theta, X^T) := \frac{dP_X}{dP_B}(X^T) = \exp\left\{ \int_0^T \frac{S(\theta, t, X(t))}{\sigma^2 X^2(t)} dX(t) - \frac{1}{2}\int_0^T \frac{S^2(\theta, t, X(t))}{\sigma^2 X^2(t)} dt \right\}, \tag{2.3.4}$$

with $S(\theta, t, X(t)) = (\varphi(t), -\ln X(t))\theta X(t)$. Then, the proof follows from classical maximization techniques. This completes the proof. □

Further, Proposition 2.3.2, the definition of $U_{[0,T]}$ in (2.3.3), and the fact that the process $\{X(t), t \geqslant 0\}$ satisfies SDE (2.2.1) imply $U_{[0,T]} = \int_0^T \frac{(\varphi(t), -\ln X(t))^\top}{X(t)} dX(t) = Q_{[0,T]}\theta +$

$\sigma W_{[0,T]}$. Then,

$$\hat{\theta}_T = Q_{[0,T]}^{-1} U_{[0,T]} = Q_{[0,T]}^{-1}\left(Q_{[0,T]}\theta + \sigma W_{[0,T]}\right) = \theta + \sigma Q_{[0,T]}^{-1} W_{[0,T]}. \tag{2.3.5}$$

This implies that $\sqrt{T}(\hat{\theta}_T - \theta) = \sigma T Q_{[0,T]}^{-1} \frac{1}{\sqrt{T}} W_{[0,T]}$. The asymptotic behavior of $\hat{\theta}_T$ relies on the matrix $TQ_{[0,T]}^{-1}$ and the column vector $\frac{1}{\sqrt{T}} W_{[0,T]}$ as T tends to infinity. Thus, below, the asymptotic behavior of the matrices $\dfrac{1}{T} Q_{[0,T]}$ and $TQ_{[0,T]}^{-1}$ is presented. Let $\tilde{B}_s = B_s \mathbb{I}_{\mathbb{R}+}(s) + \bar{B}_{-s}\mathbb{I}_{\mathbb{R}-}(s)$ be a bilateral Brownian motion, where $\{B_s\}_{s\geqslant 0}$ and $\{\bar{B}_{-s}\}_{s\geqslant 0}$ are two independent Brownian motions. The auxiliary process $\{\tilde{X}(t), t \geqslant 0\}$ is introduced in (A.58) in the Appendix A.2: $\ln \tilde{X}(t) = \tilde{r}(t) + \tilde{\tau}(t)$, where $\tilde{r}(t) = e^{-\alpha t} \int_{-\infty}^{t} e^{\alpha s}\left(L(s) - \frac{1}{2}\sigma^2\right) ds$, $\tilde{\tau}(t) = \sigma e^{-\alpha t} \int_{-\infty}^{t} e^{\alpha s} d\tilde{B}_s$. More propositions and the relation between the auxiliary process and the process $\{X(t), t \geqslant 0\}$ are given in the Appendix A.2. To establish this result, first, the invertibility of the following matrix is given. Let

$$\Sigma = \begin{bmatrix} I_p & -\displaystyle\int_0^1 \varphi^\top(t)\tilde{r}(t)dt \\ -\displaystyle\int_0^1 \varphi(t)\tilde{r}(t)dt & \displaystyle\int_0^1 (\tilde{r}(t))^2 dt + \frac{\sigma^2}{2\alpha} \end{bmatrix}, \tag{2.3.6}$$

the following result shows that Σ is a positive definite matrix.

Proposition 2.3.3. *The matrix Σ is a $(p+1)\times(p+1)$-positive definite matrix.*

The proof follows from Proposition A.17 given in Appendix A.3. The following proposition is established from Proposition 2.3.3. It is useful in deriving the strong consistency of the UMLE as well as the joint asymptotic normality of the UMLE and RMLE.

Proposition 2.3.4. *If Assumption 2.1-2.3 hold, then*

$$\frac{1}{T} Q_{[0,T]} \xrightarrow[T\to\infty]{a.s.\text{ and } L^{m/2}} \Sigma \text{ and } TQ_{[0,T]}^{-1} \xrightarrow[T\to\infty]{a.s.} \Sigma^{-1}.$$

The proof of this proposition is given in Appendix A.3. The following proposition is derived by using the martingale central limit theorem for diffusion processes. It gives the limiting distribution of $\frac{1}{\sqrt{T}} W_{[0,T]}$.

Proposition 2.3.5. *If Assumption 2.1-2.3 hold, then,* $\frac{1}{\sqrt{T}} W_{[0,T]} \xrightarrow[T\to\infty]{D} W^* \sim \mathcal{N}_{p+1}(0, \Sigma)$.

The proof of this proposition is outlined in Appendix A.3. The following proposition proves that UMLE is strongly consistent and asymptotically normal, and it follows from Proposition 2.3.4 and Proposition 2.3.5. For the sake of simplicity, let $\rho_T = \sqrt{T}(\hat{\theta}_T - \theta)$.

Proposition 2.3.6. *Suppose that Assumption 2.1-2.3 hold. Then,* $\hat{\theta}_T$ *is a strongly consistent estimator of* θ*. Furthermore,* $\hat{\theta}_T$ *is asymptotically normal, i.e.* $\rho_T \xrightarrow[T\to\infty]{D} \rho \sim \mathcal{N}_{p+1}(0, \sigma^2\Sigma^{-1})$.

The proof of this proposition is given in Appendix A.3. The asymptotic normality of the RMLE is derived from Proposition 2.3.6. More generally, the joint asymptotic normality between the UMLE and the RMLE is also derived. To this end, the following set of local alternative restrictions is considered:

$$H_{a,T} : M\theta - r = \frac{r_0}{\sqrt{T}}, \quad T > 0 \tag{2.3.7}$$

where r_0 is a fixed q-column vector. To derive the RMLE, the Lagrange multiplier λ into the log-likelihood function is introduced,

$$\log \mathcal{L}(\theta, X^T, \lambda) = \frac{1}{\sigma^2}\theta^\top U_{[0,T]} - \frac{1}{2\sigma^2}\theta^\top Q_{[0,T]}\theta + \lambda^\top(M\theta - r),$$

where λ is a q column vector. After some algebraic computation,

$$\tilde{\theta}_T = \hat{\theta}_T + G_{[0,T]}r - G_{[0,T]}M\hat{\theta}_T = \hat{\theta}_T - G_{[0,T]}(M\hat{\theta}_T - r),$$

where $G_{[0,T]} = Q_{[0,T]}^{-1}M^\top(MQ_{[0,T]}^{-1}M^\top)^{-1}$. Note that

$$\begin{aligned}\sqrt{T}(\tilde{\theta}_T - \theta) &= \sqrt{T}\left(\hat{\theta}_T - G_{[0,T]}(M\hat{\theta}_T - r) - \theta\right) \\ &= \sqrt{T}\left(I_{(p+1)} - G_{[0,T]}M\right)\hat{\theta}_T + \sqrt{T}\left(G_{[0,T]}r - \theta\right).\end{aligned}$$

This gives

$$\sqrt{T}(\tilde{\theta}_T - \theta) = \left(I_{(p+1)} - G_{[0,T]}M\right)\sqrt{T}(\hat{\theta}_T - \theta) - \sqrt{T}G_{[0,T]}\left(M\theta - r\right). \tag{2.3.8}$$

A continuous function $f(X) = XM^\top(MXM^\top)^{-1}$ is constructed, where X is a positive definite matrix. Then,

$$f(TQ_{[0,T]}^{-1}) = TQ_{[0,T]}^{-1}M^\top(MTQ_{[0,T]}^{-1}M^\top)^{-1} = G_{[0,T]}.$$

By Proposition 2.3.4 and continuous mapping theorem,

$$G_{[0,T]} \xrightarrow[T\to\infty]{\text{P}} G^* = \Sigma^{-1}M^\top(M\Sigma^{-1}M^\top)^{-1}, \tag{2.3.9}$$

and $I_{(p+1)} - G_{[0,T]}M \xrightarrow[T\to\infty]{\text{P}} I_{(p+1)} - G^*M$. Consider the local alternatives restriction (2.3.7),

$$\sqrt{T}G_{[0,T]}(M\theta - r) = \sqrt{T}G_{[0,T]}\frac{r_0}{\sqrt{T}} = G_{[0,T]}r_0 \xrightarrow[T\to\infty]{\text{P}} G^*r_0. \tag{2.3.10}$$

To simplify some notations, let $(\rho_T^\top, \varrho_T^\top, \varsigma_T^\top)^\top = \sqrt{T}\left((\hat{\theta}_T - \theta)^\top, (\tilde{\theta}_T - \theta)^\top, (\hat{\theta}_T - \tilde{\theta}_T)^\top\right)^\top$.

Proposition 2.3.7. *If Assumption 2.1-2.3 and the local alternative restriction* (2.3.7) *hold, then,* $(\rho_T^\top, \varrho_T^\top, \varsigma_T^\top)^\top \xrightarrow[T\to\infty]{D} (\rho^\top, \varrho^\top, \varsigma^\top)^\top$ *where*

$$\begin{bmatrix}\rho \\ \varrho \\ \varsigma\end{bmatrix} \sim \mathcal{N}_{3(p+1)}\left(\begin{bmatrix}0 \\ -G^*r_0 \\ G^*r_0\end{bmatrix}, \sigma^2\begin{bmatrix}\Sigma^{-1} & \Sigma^{-1} - G^*M\Sigma^{-1} & G^*M\Sigma^{-1} \\ \Sigma^{-1} - G^*M\Sigma^{-1} & \Sigma^{-1} - G^*M\Sigma^{-1} & 0 \\ G^*M\Sigma^{-1} & 0 & G^*M\Sigma^{-1}\end{bmatrix}\right).$$

The proof is provided in Appendix A.3. Proposition 2.3.7 constitutes the main result of this subsection and it is used, in the next subsection, in deriving a test for the testing problem in (2.2.3).

2.3.1.2 Testing the restriction

This subsection tackles the hypothesis testing problem in (2.2.3). Note that, in the continuous time observations, the diffusion parameter can be consistently estimated by the discretized version of quadratic variation of the process $\{\ln X(t), t \geqslant 0\}$. This is given by $\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{n}(\ln X(t_i) - \ln X(t_{i-1}))^2$, with $0 = t_0 < t_1 < \cdots < t_n < T$. It is well known that $\hat{\sigma}^2$ is the consistent estimator of the diffusion coefficient σ^2 for $T > 0$ and $\max_{1\leqslant i\leqslant n}\{t_i - t_{i-1}\} \to 0$. To introduce some notations, let $\chi_q^2(\lambda)$ be the chi-square random variable with q degrees of freedom and non-centrality parameter $\lambda \geqslant 0$. In particular, if $\lambda = 0$, χ^2_q is used to stand for a (central) chi-square random variable, with q degrees of freedom. Further, let $\chi^2_{\alpha;q}$ be the α^{th} quantile of a χ^2_q where $0 < \alpha \leqslant 1$, let $\Delta = \frac{1}{\sigma^2} r_0^\top (M\Sigma^{-1}M^\top)^{-1} r_0$, with r_0 given in (2.3.7), and let $\varsigma_T = \sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T)$, $\hat{\Gamma} = \frac{1}{\hat{\sigma}^2} M^\top \left(MTQ^{-1}_{[0,T]} M^\top \right)^{-1} M$, $\Gamma = \frac{1}{\sigma^2} M^\top \left(M\Sigma^{-1}M^\top \right)^{-1} M$, $\psi_T = \varsigma_T^\top \hat{\Gamma} \varsigma_T$, $\psi = \varsigma^\top \Gamma \varsigma$, $\psi_0 = \varsigma_0^\top \Gamma \varsigma_0$ where $\varsigma_0 \sim \mathcal{N}_{p+1}\left(0, \sigma^2 G^* M\Sigma^{-1}\right)$. The following proposition gives the asymptotic distribution of the test statistics ψ_T.

Proposition 2.3.8. *If Assumption 2.1-2.3 hold, then, if* $r_0 \neq 0$, *then,* $\psi_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi \sim \chi^2_q(\Delta)$; *if* $r_0 = 0$, *then,* $\psi_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi_0 \sim \chi^2_q$.

The proof of this proposition is provided Appendix A.3. From Proposition 2.3.8, to test the null hypothesis in (2.2.3) is suggested by using the rejection region $\psi_T > \chi^2_{\alpha;q}$ for a given α. Thus, the suggested test is

$$\kappa_T = \mathbb{I}_{\{\psi_T > \chi^2_{\alpha;q}\}} \tag{2.3.11}$$

From Proposition 2.3.8, below, the asymptotic power of the proposed test is derived.

Proposition 2.3.9. *Suppose that Assumption 2.1-2.3 hold, along with local alternative restriction (2.3.7). Then, the asymptotic power function of the test* κ_T *in* (2.3.11) *is given by* $\Pi(\Delta) = \mathrm{P}\left(\chi^2_q(\Delta) \geqslant \chi^2_{\alpha;q}\right)$.

The proof of this Proposition follows directly from Proposition 2.3.8. It should be noted that, under the null hypothesis, $\Delta = 0$, the above asymptotic local power is equal to α. Further, if Δ tends to infinity, the above asymptotic local power tends to 1. The numerical results show that asymptotic local power is increasing to 1 as time T increases to infinity. For more details about the inference under the setting of no change-point, we refer to Lyu and Nkurunziza [2023c].

2.3.2 The case of a possible change point

Over the years, the problem of a change-point detection in a stochastic process has been an important issue in statistical inference. Initially investigated for i.i.d. data, change-point analysis has been more recently extended to time series of dependent data. This section investigates the problem of detecting changes in the drift parameter of the GEOU process in (2.2.1). It is interesting to test whether there is a change in the component of the drift in the time interval $[0, T]$, during which the process is observed. Estimating the change-point is also an interesting problem. This chapter considers that the change-point is $t^* = sT$ with $s \in (0, 1)$, and thus, the GEOU process with change-point is given by

$$dX(t) = \left(S(\theta^{(1)}, t, X(t))\mathbb{I}_{\{t \leq t^*\}} + S(\theta^{(2)}, t, X(t))\mathbb{I}_{\{t > t^*\}}\right) dt + \sigma X(t) dB_t, \; X(0) = X_0 \quad (2.3.12)$$

where for $j = 1, 2$,

$$S(\theta^{(j)}, t, X(t)) = (L(t) - \alpha_j \ln X(t))X(t) = \left(\sum_{i=1}^{p} \mu_i \varphi_i(t) - \alpha_j \ln X(t)\right) X(t). \quad (2.3.13)$$

Let $\theta^{(1)} = (\mu_{11}, \mu_{12}, \ldots, \mu_{1p}, \alpha_1)^\top$, $\theta^{(2)} = (\mu_{21}, \mu_{22}, \ldots, \mu_{2p}, \alpha_2)^\top$, $\boldsymbol{\theta} = (\theta^{(1)\top}, \theta^{(2)\top})^\top$. As in (2.2.2), the case where the parameter $\boldsymbol{\theta}$ may satisfy the restriction is also considered:

$$\tilde{H}_0 : \tilde{M}\boldsymbol{\theta} = \tilde{r} \quad (2.3.14)$$

where, for a fixed p, $\tilde{M}$ is a known $q \times 2(p + 1)$ full rank matrix with $q < 2(p + 1)$, $\tilde{r}$ is a known q column vector. This restriction leads to the hypothesis testing problem

$$\tilde{H}_0 : \tilde{M}\boldsymbol{\theta} = \tilde{r} \quad \text{verus} \quad H_1 : \tilde{M}\boldsymbol{\theta} \neq \tilde{r}. \quad (2.3.15)$$

Note that if $\tilde{M} = [I_{p+1}, -I_{p+1}]$ and $\tilde{r} = 0$, the testing problem is for detecting the existence of the change-point, i.e. detecting the existence of the change-point is a particular case of the hypothesis (2.3.15). Thus, inference methods to be established are similar to that given in Section 2.3.1. As in Section 2.3.1, the following proposition gives the expression of the process $\{X(t), t \geqslant 0\}$ as the solution to the SDE (2.3.12).

Proposition 2.3.10. *Suppose that Assumption 2.1-2.3 hold. Then, the solution of SDE (2.3.12) is given*

$$X(t) = \exp\{\ln X_1(t)\mathbb{I}_{\{t\leqslant t^*\}} + \ln X_2(t)\mathbb{I}_{\{t>t^*\}}\}, \qquad \ln X_1(t) = e^{-\alpha_1 t}\ln X_0 + r_1(t) + \tau_1(t),$$
$$\ln X_2(t) = e^{-\alpha_2(t-t^*)}\ln X^{t^*}(0) + r_2^{t^*}(t-t^*) + \tau_2^{t^*}(t-t^*), \tag{2.3.16}$$

where for $0 \leqslant t \leqslant t^*$, $r_1(t) = e^{-\alpha_1 t}\int_0^t e^{\alpha_1 s}\left(L(s) - \frac{1}{2}\sigma^2\right)ds$, $\tau_1(t) = \sigma e^{-\alpha_1 t}\int_0^t e^{\alpha_1 s}dB_s$ *and for* $t^* \leqslant t \leqslant T$, $r_2^{t^*}(t) = e^{-\alpha_2 t}\int_0^t e^{\alpha_2 s}\left(L^{t^*}(s) - \frac{1}{2}\sigma^2\right)ds$, $\tau_2^{t^*}(t) = \sigma e^{-\alpha_2 t}\int_0^t e^{\alpha_2 s}dB_s^{t^*}$, $\ln X^{t^*}(t) = \ln X(t+t^*)$, $L^{t^*}(t) = L(t+t^*)$, $B_t^{t^*} = B(t+t^*) - B(t^*)$, $t \geqslant 0$ $t^* \geqslant 0$.

Proof. The proof is similar to the proof of Proposition 2.2.1. □

The following result is useful in deriving the likelihood function of the GEOU process in (2.3.12) in the context where the change-point is known. It is derived from Proposition 2.3.10. Indeed, as intermediate step, the change-point $t^* = sT$ is supposed to be known, i.e. s is known.

Proposition 2.3.11. *If Assumption 2.1-2.3, then*

(1) $\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m] < \infty$ *(2)* $\sup_{t\geqslant 0}\mathbb{E}[X(t)] < \infty$, *(3)* $\sup_{t\geqslant 0}\mathbb{E}[X^m(t)] < \infty$,

(4) $\sup_{t\geqslant 0}\mathrm{E}\left(\left(\left|S(\theta^{(1)}, t, X(t))\right|^m + \left|S(\theta^{(2)}, t, X(t))\right|^m\right)\Big/|\sigma X(t)|^m\right) < \infty$,

(5) $\mathrm{P}\left(\int_0^T \left(\left|S(\theta^{(1)}, t, X(t))\right|^m + \left|S(\theta^{(2)}, t, X(t))\right|^m\right)\Big/|\sigma X(t)|^m\, dt < \infty\right) = 1$, $\forall\ 0 \leqslant T < \infty$.

The proof of this proposition is given in Appendix A.3. From Proposition 2.3.11, one concludes that the Radon-Nikodym derivative of the SDE in (2.3.12) exists and it is given by

$$\mathcal{L}(\boldsymbol{\theta}, s, X^T) := \frac{dP_X}{dP_B}(X^T) = \exp\left\{\int_0^T \frac{S(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)} \mathbb{I}_{\{0\leqslant t\leqslant t^*\}} dX(t)\right.$$
$$+ \int_0^T \frac{S(\boldsymbol{\theta},t,X(t))}{\sigma^2 X^2(t)} \mathbb{I}_{\{t^*<t\leqslant T\}} dX(t) \tag{2.3.17}$$
$$\left.-\frac{1}{2}\int_0^T \frac{S^2(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)} \mathbb{I}_{\{0\leqslant t\leqslant t^*\}} dt - \frac{1}{2}\int_0^T \frac{S^2(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)} \mathbb{I}_{\{t^*<t\leqslant T\}} dt\right\}.$$

Then, the UMLE is derived by maximizing the functional $\boldsymbol{\theta} \longmapsto \mathcal{L}(\boldsymbol{\theta}, X^T)$. Further, RMLE is derived by using Lagrange multiplier method. Let $\hat{\theta}_T(s)$ be the UMLE and $\tilde{\theta}_T(s)$ be the RMLE. Note that the process $\{X(t), t \geqslant 0\}$ is not stationary. Because of that, to study the long term behavior of proposed estimators for the case of a known change-point, an auxiliary process is constructed. The introduced process is *close in certain sense* to the solution of the SDE (2.3.12). In particular, let $\tilde{B}_s$ is a bilateral Brownian motion. i.e. $\tilde{B}_s = B_s \mathbb{I}_{\mathbb{R}^+}(s) + \bar{B}_{-s}\mathbb{I}_{\mathbb{R}^-}(s)$, where $\{B_s\}_{s\geqslant 0}$ and $\{\bar{B}_{-s}\}_{s\geqslant 0}$ are two independent Brownian motions.A new process $\{\tilde{X}_t, t \geqslant 0\}$ isintroduced, where

$$\tilde{X}(t) = \exp\left\{\ln \tilde{X}_1(t)\mathbb{I}_{\{t\leqslant t^*\}} + \ln \tilde{X}_2(t)\mathbb{I}_{\{t>t^*\}}\right\}, \qquad \ln \tilde{X}_1(t) = e^{-\alpha_1 t}\ln X_0 + \tilde{r}_1(t) + \tilde{\tau}_1(t),$$
$$\ln \tilde{X}_2(t) = e^{-\alpha_2(t-t^*)}\ln \tilde{X}^{t^*}(0) + \tilde{r}_2^{t^*}(t-t^*) + \tilde{\tau}_2^{t^*}(t-t^*), \tag{2.3.18}$$

where for $0 \leqslant t \leqslant t^*$, $\tilde{r}_1(t) = e^{-\alpha_1 t}\int_{-\infty}^t e^{\alpha_1 s}\left(L(s) - \frac{1}{2}\sigma^2\right)ds$, $\tilde{\tau}_1(t) = \sigma e^{-\alpha_1 t}\int_{-\infty}^t e^{\alpha_1 s} d\tilde{B}_s$ and for $t^* \leqslant t \leqslant T$, $\tilde{r}_2^{t^*}(t) = e^{-\alpha_2 t}\int_{-\infty}^t e^{\alpha_2 s}\left(L^{t^*}(s) - \frac{1}{2}\sigma^2\right)ds$, $\tilde{\tau}_2^{t^*}(t) = \sigma e^{-\alpha_2 t}\int_{-\infty}^t e^{\alpha_2 s} d\tilde{B}_s^{t^*}$, $\ln \tilde{X}^{t^*}(t) = \ln \tilde{X}(t+t^*)$, $L^{t^*}(t) = L(t+t^*)$, $\tilde{B}_t^{t^*} = \tilde{B}(t+t^*) - \tilde{B}(t^*)$, $t \geqslant 0$ $t^* \geqslant 0$. First, the proof of that the sequence of random variables $\{\ln \tilde{X}_k(t+j-1)\}_{j=1}^{\infty}$ is stationary and ergodic is given. Further, it is proven that the distance between $\ln X(t)$ and $\ln \tilde{X}(t)$ converges, almost surely and in mean, to 0 as t tends to infinity.

Proposition 2.3.12. *Suppose that Assumption 2.1-2.3 hold. Then, for* $t \in [0,1]$*, and* $k = 1, 2$, $j = 1, 2, \cdots$*, the sequence of random variables* $\{\ln \tilde{X}_k(t+j-1)\}_{j=1}^{\infty}$ *is stationary and ergodic.*

The proof is similar to the proof of Proposition A.15 in Appendix A.3.

Proposition 2.3.13. *If Assumption 2.1-2.3 hold, then,*

(1) $|\ln \tilde{X}_k(t) - \ln X_k(t)| \xrightarrow[T\to\infty]{a.s. \text{ and } L^m} 0$, *for* $k = 1, 2$,

(2) $\left(\frac{1}{T}\int_0^{sT}\left(\varphi^\top(t)\ln X_1(t) - \varphi^\top(t)\ln \tilde{X}_1(t)\right)dt, \frac{1}{T}\int_{sT}^{T}\left(\varphi^\top(t)\ln X_2(t) - \varphi^\top(t)\ln \tilde{X}_2(t)\right)dt\right) \xrightarrow[T\to\infty]{a.s. \text{ and } L^m} \mathbf{0}$,

(3) $\left(\frac{1}{T}\int_0^{sT}\left((\ln X_1(t))^2 - (\ln \tilde{X}_1(t))^2\right)dt, \frac{1}{T}\int_{sT}^{T}\left((\ln X_2(t))^2 - (\ln \tilde{X}_2(t))^2\right)dt\right) \xrightarrow[T\to\infty]{a.s. \text{ and } L^{m/2}} \mathbf{0}$.

Proof. The proof of the first claim is similar to Proposition A.16 in Appendix A.2. The second and third statements follow from the first claim along with the continuous version of Cesàro mean theorem. □

Further, let $\hat{\theta}_T(s)$ be the UMLE and $\tilde{\theta}_T(s)$ be the RMLE in the context of a known change point. RMLE is derived by using Lagrange multiplier method. Proposition 2.3.13 along with the ergodicity of the auxiliary process are useful in deriving the asymptotic distributions of $\hat{\theta}_T(s)$ and $\tilde{\theta}_T(s)$.

2.3.2.1 Inference in the case where there is a known change-point

This subsection considers inference problem about $\boldsymbol{\theta}$ in the context of known rate of the change-point location s. To simplify some mathematical expressions, some new notations are introduced. Let

$$Q(s,T) = \begin{bmatrix} Q_{[0,sT]} & 0 \\ 0 & Q_{[sT,T]} \end{bmatrix}, U(s,T) = \begin{bmatrix} U_{[0,sT]} \\ U_{[sT,T]} \end{bmatrix}, \qquad W(s,T) = \begin{bmatrix} W_{[0,sT]} \\ W_{[sT,T]} \end{bmatrix}. \quad (2.3.19)$$

The following proposition gives the sufficient condition for the existence of UMLE and RMLE.

Proposition 2.3.14. *Suppose that Assumption 2.1-2.3 hold. Then, the matrix* $Q_{[0,sT]}$, $Q_{[sT,T]}$, *and* $Q(s,T)$ *are positive definite, provided that* $\min\{sT, (1-s)T\} \geqslant 1$.

Proof. Similarly to the proof of Proposition 2.3.1, it is proven that the matrices $Q_{[0,sT]}$, $Q_{[sT,T]}$ are positive definite, under the condition that $sT \geqslant 1$, $(1-s)T \geqslant 1$, respectively. This implies that the matrix $Q(s,T)$ is a positive definite provided that $\min\{sT,(1-s)T\} \geqslant 1$. □

As mentioned in Section 2.3.1, the condition $\min\{sT,(1-s)T\} \geqslant 1$ will be considered as satisfied because the proposed method is asymptotic.

Let $\tilde{G}_{[0,T]} = Q^{-1}(s,T)\tilde{M}^{\top}(\tilde{M}Q^{-1}(s,T)\tilde{M}^{\top})^{-1}$.

Proposition 2.3.15. *Suppose that Assumption 2.1-2.3 hold along with (2.3.15). Then,* $\hat{\theta}_T(s) = Q(s,T)^{-1}U(s,T)$ *and* $\tilde{\theta}_T(s) = \hat{\theta}_T(s) - \tilde{G}_{[0,T]}(\tilde{M}\hat{\theta}_T(s) - \tilde{r})$.

The proof of this proposition is given in Appendix A.3.

2.3.2.2 Joint asymptotic normality of the estimators

This subsection derives the joint asymptotic normality of UMLE $\hat{\theta}_T(s)$ and RMLE $\tilde{\theta}_T(s)$. This result is used in deriving a test for the hypothesis testing problem (2.3.15). As in Section 2.3.1, from Proposition 2.3.15, $\sqrt{T}(\hat{\theta}_T(s) - \boldsymbol{\theta}) = \sigma T Q^{-1}(s,T)\frac{1}{\sqrt{T}}W(s,T)$. Thus, as in Section 2.3.1, the asymptotic properties of these estimators relies on the asymptotic behavior of $\frac{1}{T}Q(s,T)$ and $\frac{1}{\sqrt{T}}W(s,T)$ as well as on the non-singularity of the matrices

$$\tilde{\Sigma} = \begin{bmatrix} s\Sigma_1 & 0 \\ 0 & (1-s)\Sigma_2 \end{bmatrix}, \tag{2.3.20}$$

where for $k = 1,2$,

$$\Sigma_k = \begin{bmatrix} I_p & -\int_0^1 b^{\top}(t)\tilde{r}_k(t)dt \\ -\int_0^1 \varphi(t)\tilde{r}_k(t)dt & \int_0^1 (\tilde{r}_k(t))^2 dt + \dfrac{\sigma^2}{2\alpha_k} \end{bmatrix}. \tag{2.3.21}$$

Proposition 2.3.16. *The matrix* $\tilde{\Sigma}$ *is a positive definite matrix.*

Proof. Similar to the proof of Proposition 2.3.3. □

Proposition 2.3.17. *If Assumption 2.1-2.3 hold, then* $\frac{1}{T}Q_{[0,sT]} \xrightarrow[T\to\infty]{a.s. \text{ and } L^{m/2}} s\Sigma_1$, $\frac{1}{T}Q_{[sT,T]} \xrightarrow[T\to\infty]{a.s. \text{ and } L^{m/2}} (1-s)\Sigma_2$, $TQ^{-1}_{[0,sT]} \xrightarrow[T\to\infty]{a.s.} \frac{1}{s}\Sigma_1^{-1}$, $TQ^{-1}_{[sT,T]} \xrightarrow[T\to\infty]{a.s.} \frac{1}{1-s}\Sigma_2^{-1}$, $\frac{1}{T}(Q(s,T)) \xrightarrow[T\to\infty]{a.s. \text{ and } L^{m/2}} \tilde{\Sigma}$, $T(Q^{-1}(s,T)) \xrightarrow[T\to\infty]{a.s.} \tilde{\Sigma}^{-1}$. *Further,* $\frac{1}{\sqrt{T}}W(s,T) \xrightarrow[T\to\infty]{\mathrm{D}} \tilde{W}^* \sim \mathcal{N}_{2(p+1)}\left(0,\tilde{\Sigma}\right)$.

Proof of Proposition 2.3.17. The first six statements are established in the similar way as in Proposition 2.3.4. The proof of last statement follows from the first six statements of the proposition along with the martingale central limit theorem for diffusion processes. □

Let $\tilde{\rho}_T = \sqrt{T}(\hat{\theta}_T(s) - \boldsymbol{\theta})$. From Proposition 2.3.17, the following proposition is derived, which shows that $\hat{\theta}_T(s)$ is consistent and asymptotically normal.

Proposition 2.3.18. *Suppose that Assumption 2.1-2.3 hold. Then,* $\hat{\theta}_T(s)$ *is a strongly consistent estimator of* $\boldsymbol{\theta}$. *Furthermore,* $\hat{\theta}_T(s)$ *is asymptotically normal, i.e.* $\tilde{\rho}_T \xrightarrow[T\to\infty]{D} \tilde{\rho} \sim \mathcal{N}_{2(p+1)}(0, \sigma^2\tilde{\Sigma}^{-1})$.

The proof of this proposition is given in Appendix A.1. From(2.3.8),

$$\sqrt{T}(\tilde{\theta}_T(s) - \boldsymbol{\theta}) = \left(I_{(p+1)} - \tilde{G}_{[0,T]}\tilde{M}\right)\sqrt{T}(\hat{\theta}_T(s) - \boldsymbol{\theta}) - \sqrt{T}\tilde{G}_{[0,T]}\left(\tilde{M}\boldsymbol{\theta} - \tilde{r}\right). \quad (2.3.22)$$

Below, the following derived proposition gives the joint asymptotic normality of UMLE and RMLE. Let $(\tilde{\rho}_T^\top, \tilde{\varrho}_T^\top, \tilde{\varsigma}_T^\top)^\top = \sqrt{T}\left((\hat{\theta}_T(s) - \boldsymbol{\theta})^\top, (\tilde{\theta}_T(s) - \boldsymbol{\theta})^\top, (\hat{\theta}_T(s) - \tilde{\theta}_T(s))^\top\right)^\top$, and let $\tilde{G}^* = \tilde{\Sigma}^{-1}\tilde{M}^\top(\tilde{M}\tilde{\Sigma}^{-1}\tilde{M}^\top)^{-1}$. Suppose the following set of local alternative restrictions hold,

$$H_{\tilde{a},T} : \tilde{M}\boldsymbol{\theta} - \tilde{r} = \frac{\tilde{r}_0}{\sqrt{T}}, T > 0 \quad (2.3.23)$$

where $\tilde{r}_0$ is a fixed q-column vector.

Proposition 2.3.19. *Suppose that Assumption 2.1-2.3 and the local alternative restriction (2.3.23) hold. Then,* $(\tilde{\rho}_T^\top, \tilde{\varrho}_T^\top, \tilde{\varsigma}_T^\top)^\top \xrightarrow[T\to\infty]{D} (\tilde{\rho}^\top, \tilde{\varrho}^\top, \tilde{\varsigma}^\top)^\top$*, where*

$$\begin{bmatrix} \tilde{\rho} \\ \tilde{\varrho} \\ \tilde{\varsigma} \end{bmatrix} \sim \mathcal{N}_{6(p+1)} \left[\begin{bmatrix} 0 \\ -\tilde{G}^* \tilde{r}_0 \\ \tilde{G}^* \tilde{r}_0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \tilde{\Sigma}^{-1} & \tilde{\Sigma}^{-1} - \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1} & \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1} \\ \tilde{\Sigma}^{-1} - \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1} & \tilde{\Sigma}^{-1} - \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1} & 0 \\ \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1} & 0 & \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1} \end{bmatrix} \right].$$

Proof. The proof is similar to Proposition 2.3.7. □

2.3.2.3 Testing the restriction

This subsection handles with the problem of testing the hypothesis in (2.3.15) within the setting of one known change-point. Let

$$\begin{aligned} \Delta &= \tfrac{1}{\sigma^2} \tilde{r}_0^\top \left(\tilde{M} \tilde{\Sigma}^{-1} \tilde{M}^\top \right)^{-1} \tilde{r}_0, \quad \Gamma = \tfrac{1}{\sigma^2} \tilde{M}^\top \left(\tilde{M} \tilde{\Sigma}^{-1} \tilde{M}^\top \right)^{-1} \tilde{M}, \\ \hat{\Gamma} &= \tfrac{1}{\hat{\sigma^2}} \tilde{M}^\top \left(\tilde{M} T Q^{-1}(s, T) \tilde{M}^\top \right)^{-1} \tilde{M}. \end{aligned} \tag{2.3.24}$$

Further, let $\tilde{\psi}_T = \tilde{\varsigma}_T^\top \hat{\Gamma} \tilde{\varsigma}_T$, $\tilde{\psi} = \tilde{\varsigma}^\top \Gamma \tilde{\varsigma}$ and $\tilde{\psi}_0 = \tilde{\varsigma}_0^\top \Gamma \tilde{\varsigma}_0$, where $\tilde{\varsigma}_0 \sim \mathcal{N}_{2(p+1)}(0, \sigma^2 \tilde{G}^* \tilde{M} \tilde{\Sigma}^{-1})$ and $\tilde{G}^* = \tilde{\Sigma}^{-1} \tilde{M}^\top \left(\tilde{M} \tilde{\Sigma}^{-1} \tilde{M}^\top \right)^{-1}$. The following proposition is derived, which is important in developing the suitable test for testing the hypothesis problem (2.3.15).

Proposition 2.3.20. *Suppose that Assumption 2.1-2.3 hold. Then, if* $\tilde{r}_0 \neq 0$, $\tilde{\psi}_T \xrightarrow[T\to\infty]{\mathrm{D}} \tilde{\psi} \sim \chi_q^2(\tilde{\Delta})$*, and if* $\tilde{r}_0 = 0$, $\tilde{\psi}_T \xrightarrow[T\to\infty]{\mathrm{D}} \tilde{\psi}_0 \sim \chi_q^2$.

Proof. The proof is similar to Proposition 3.4.1 □

In this case, the null hypothesis in (2.3.15) is tested by using the rejection region $\tilde{\psi}_T > \chi^2_{\alpha;q}$ for a given α. The suggested the test is

$$\tilde{\kappa}_T = \mathbb{I}_{\{\tilde{\psi}_T > \chi^2_{\alpha;q}\}}. \tag{2.3.25}$$

From Proposition 2.3.20, below, the asymptotic power function of the proposed test is derived.

Proposition 2.3.21. *Suppose that Assumption 2.1-2.3 along with (2.3.23) hold. Then, the asymptotic power function of the test $\tilde{\kappa}_T$ in (2.3.25) is given by $\widetilde{\varpi}(\Delta) = \mathrm{P}\left(\chi^2_q(\Delta) \geqslant \chi^2_{\alpha;q}\right)$.*

Proof. The proof follows directly from Proposition 2.3.20. □

Similar to Proposition 3.4.2, under the null hypothesis, $\Delta = 0$, the above asymptotic local power is equal to α. Further, if Δ tends to infinity, the above asymptotic local power tends to 1 and our numerical results show that asymptotic local power is also increasing to 1 as time T increases to infinity. For more details about the inference under the setting of one potential change-point, we refer to Lyu and Nkurunziza [2023d].

2.4 Inference in the case where p is unknown

2.4.1 The case of absent change-point

In most related references, such as Dehling et al. [2014], Nkurunziza and Zhang [2018], among others, the dimension of the base functions, p is supposed to be known. But in practice, p is not known and thus, it is important to develop a statistical method for estimating the appropriate number of the base functions, i.e. p. This subsection considers a more general inference problem about the drift parameter θ when the nuisance parameter p is unknown and needs to be estimated. Detecting the suitable value of p corresponds to solving a model selection problem. In other words, selecting the best statistical model according to the log-likelihood-based information criterion from all candidate models. To decide the reasonable value of p, the following *Schwartz Information Criterion method* is used. This method consists in minimising the log-likelihood-based information criterion $\mathrm{IC}(p) = -2\log\mathcal{L}([0,T], \hat{\theta}_T) + h(p)\Phi(T)$, where the log-likelihood function $\log\mathcal{L}([0,T], \hat{\theta}_T)$ is defined in (2.3.1), $h(p) = p + 1$ is the number of drift parameters, p is the potential number of the base functions and $\Phi(T)$ is a non-decreasing function of T.

As known, in statistical modeling, AIC and SIC (BIC) are two important information criteria for model selection among a collection of viable candidate models based on the likelihood function. SIC criterion is first derived by Schwarz [1978] and it has a close relation to AIC. When fitting the models, if the number of parameters is added, the value of the likelihood function will increase, but doing this may lead to overfitting. A penalty term for the number of parameters in the model is introduced to avoid overfitting. In our case, $h(p)\Phi(T)$ is the penalty term. Asymptotically, the criterion of SIC is consistent in the case where the true model is considered among the candidate models. Further, if the true model is not among the candidate models, the criterion AIC is more efficient [Vrieze, 2012]. Because in our case, under the conditions of hypothesis, the true model is one of the candidates, SIC criterion is chosen. To this end, a discretized version of the GEOU process in (2.2.1) is considered. In applied mathematics, discretization is the process of transferring continuous functions, models, variables and equations into discrete counterparts. This process is usually carried out as a first step toward making them suitable for numerical evaluation and implementation on computers. Usually, in practice the data are observed in discrete time, which implies that the integrals $\int_0^T S(\theta, t, X(t))dX(t)$ and $\int_0^T S^2(\theta, t, X(t))dt$ can be approximated using appropriate finite sums that depend on some discrete sampling for which the step of discretisations are small. Without loss of generality, Euler-Maruyama discretization, with the partition $0 = t_0 < t_1 < \cdots < t_N = T$ on a given period $[0, T]$ with $\Delta_N = \max\limits_{1 \leqslant i \leqslant N}(t_{i+1} - t_i)$ is considered. The adopted discretization scheme corresponds to the scenario of high-frequency data with a large-observation horizon i.e. the scenario where T tends to infinity while the discretization step tends to zero is considered. For the sake of simplicity, let $\mathbb{N}[a, b] = \{i \in \{0, 1, 2, \cdots, N-1\} : t_i \in [a, b]\}$. In particular, to perform the work, the following assumption is needed.

Assumption 2.4. For the partition $0 = t_0 < t_1 < \cdots < t_N = T$ on a given period $[0, T]$, there exists a constant $\boldsymbol{a}$, $\boldsymbol{a} > 1$, such that $N = O(T^{\boldsymbol{a}})$ and all the subintervals $t_{i+1} - t_i$ are supposed to be equal, for $i \in \mathbb{N}[0, T]$, i.e. $\Delta_N = T/N$.

(2.2.1) and together with Euler-Maruyama discretization imply that

$$X(t_{j+1}) - X(t_j) = \left(\sum_{i=1}^{p} \mu_i \varphi_i(t_j) - \alpha \ln X(t_j)\right) X(t_j) \left(t_{j+1} - t_j\right) + \sigma X(t_j)(B_{t_{j+1}} - B_{t_j}), \quad (2.4.1)$$

with $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$. The relation (2.4.1) gives $\left(X(t_{j+1}) - X(t_j)\right) / X(t_j) = \left(\sum_{i=1}^{p} \mu_i \varphi_i(t_j) - \alpha \ln X(t_j)\right) \left(t_{j+1} - t_j\right) + \sigma(B_{t_{j+1}} - B_{t_j})$, this can be rewritten as

$$\left(X(t_{j+1}) - X(t_j)\right) \Big/ X(t_j) = (\varphi_1(t_j), \varphi_2(t_j), \ldots, \varphi_p(t_j), - \ln X(t_j)) \left(t_{j+1} - t_j\right) \theta(p) + \sigma(B_{t_{j+1}} - B_{t_j}),$$

where $\theta(p) = \theta$ with dimension p, $0 = t_0 < t_1 < t_2 < \cdots < t_N = T$. Let $Y_i = (X(t_{i+1}) - X(t_i))/X(t_i)$, for $i \in \mathbb{N}[0, T]$, and let

$$Z_i(p) = (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_p(t_i), - \ln X(t_i))(t_{i+1} - t_i), \quad i \in \mathbb{N}[0, T]. \quad (2.4.2)$$

Then, from the discretized version of the process in (2.4.1),

$$Y_i = Z_i(p)\theta(p) + \epsilon_i, \quad i \in \mathbb{N}[0, T], \quad (2.4.3)$$

with

$$\epsilon_i = \sigma(B_{t_{i+1}} - B_{t_i}), \quad i \in \mathbb{N}[0, T]. \quad (2.4.4)$$

Here ϵ_i, $i = 1, 2, \ldots$ is the error term. Note that $\epsilon_1, \epsilon_2, \ldots$ are independent with $\epsilon_i \sim \mathcal{N}(0, \sigma^2(t_{i+1} - t_i))$, $i = 1, 2, \ldots$. In passing, let us first note a relationship between the complete sample $\{X(t) : 0 \leqslant t \leqslant T\}$ and $\left\{X^{\Delta_N}(t_i) : i = 0, 1, 2, \ldots\right\}$ where $\ln(X^{\Delta_N}(t_i))$ satisfies the Euler-Maruyama discretized version of the SDE of $\ln(X(t))$. To this end, let $i_t = \max\{i = 0, 1, 2, \cdots : t_i \leqslant t\}$ i.e. i_t is the maximum of integer less than or equal to t, and suppose that $X^{\Delta_N}(0) = X_0$. From Theorem 9.6.2 in [Kloeden and Platen, 1999, Page

324], there exists a constant C, such that

$$\sup_{0\leqslant t\leqslant T}\mathbb{E}\left[\left|\ln X(t)-\ln X^{\Delta_N}(i_t)\right|\right]\leqslant C\sqrt{\Delta_N}.$$

Let $\Psi(t)=(\varphi(t),-\ln X(t))$, for $T>0$, the following proposition is useful in proving that the estimators obtained from (2.4.3) is also consistent.

Proposition 2.4.1. *Suppose that Assumption 2.1-2.3, and Assumption 2.4 hold. Then, for $T>0$,*

(1) $\mathbb{E}\left[\left\|\sum_{k=0}^{N-1}\Psi^{\top}(t_k)\Psi(t_k)(t_{k+1}-t_k)-\int_0^T\Psi(t)^{\top}\Psi(t)dt\right\|^{m/2}\right]\leqslant K(m,\Delta_N)O(T^{m/2}),$

(2) $\mathbb{E}\left[\left\|\sum_{k=0}^{N-1}\Psi(t_k)(B_{t_{k+1}}-B_{t_k})-\int_0^T\Psi(t)dB_t\right\|^{m}\right]$
$\leqslant C_{m/2}\max\left\{(\Delta_N)^{m/2-1}3^{m-1}\left((\alpha\Delta_N)^{m/2}+o((\Delta_N)^{m/2})\right),(C_3(\Delta_N))^m(\Delta_N)^{m/2-1}\right\}T$

where

$$\begin{aligned}K(m,\Delta_N)=\max\Big\{&(C_1(\Delta_N))^{m/2},\\&\sqrt{2^{m-1}\left(K_\varphi^m3^{m-1}\left(f(m,\Delta_N)+o\left((\Delta_N)^m\right)\right)+\sup_{t\geqslant0}\mathbb{E}[|\ln X(t)|^m]\left(C_2(\Delta_N)\right)^m\right)},\\&2^{m/2-1}\sqrt{\sup_{t\geqslant0}\mathbb{E}[|\ln X(t)|^m]}\sqrt{3^{m-1}(f(m,\Delta_N)+o\left((\Delta_N)^m\right))}\Big\},\end{aligned}$$

and

$$\begin{aligned}f(m,\Delta_N)&=\left(\mathbb{E}\left[|\ln X_0|^m\right](\alpha)^m+\left(K_\mu K_\varphi\right)^m3^m\right)(\Delta_N)^m\\&\quad+\sigma^m2^{m-1}C_{m/2}\left((2\alpha)^m(\Delta_N)^m+(\alpha)^{m/2}(\Delta_N)^{m/2}\right),\end{aligned}$$

with $C_i(\Delta_N)$ is a non decreasing function with respect of Δ_N, and $\inf_{\Delta_N>0}C_i(\Delta_N)=0$ for $i=1,2,3$.

The proof of this proposition is given in Appendix A.3. In the following, let $Q_{[0,T]}(p)$ be the matrix $Q_{[0,T]}$ defined in (3.3.1) and $W_{[0,T]}(p)$ to denote the column vector $W_{[0,T]}$

defined in (2.3.3) with dimension $p+1$. Proposition 2.4.1 implies that

$$\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1}-t_i}-\frac{1}{T}Q_{[0,T]}(p)\right\|+\left\|\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,T]}\frac{\epsilon_i Z_i(p)}{t_{i+1}-t_i}-\frac{\sigma}{\sqrt{T}}W_{[0,T]}(p)\right\|\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}}0. \tag{2.4.5}$$

These formula constitutes a bridge between the discrete and continuous parameter estimation, and it plays an important role in proving the main result of this chapter by using the Schwartz Information Criterion. As in Le Breton [1976], one can approximate $\log\mathcal{L}([0,T],\theta)$ by

$$\log\mathcal{L}_N([0,T],\theta)=\frac{1}{\sigma^2}\sum_{k=0}^{N-1}\frac{S(\theta,t_k,X(t_k))}{X^2(t_k)}(X(t_{k+1})-X(t_k))-\frac{1}{2\sigma^2}\sum_{k=0}^{N-1}\frac{S^2(\theta,t_k,X(t_k))}{X^2(t_k)}(t_{k+1}-t_k).$$

Proposition 2.4.1 is useful in proving that $\log\mathcal{L}_N([0,T],\theta)$ is a good approximation for $\log\mathcal{L}([0,T],\theta)$. This is established in the following corollary.

Corollary 2.4.1. *Suppose Θ_0 is a compact subset of the parameter space Θ. Then,*

$$\mathbb{E}\left[\|\log\mathcal{L}_N([0,T],\theta)-\log\mathcal{L}([0,T],\theta)\|^{m/2}\right]\leqslant 2^{m/2-1}M_0K(m,\Delta_N)O(T^{m/2})+2^{m/2-1}M_0\times$$

$$\sqrt{C_{m/2}\max\left\{(\Delta_N)^{m/2-1}3^{m-1}\left((\alpha\Delta_N)^{m/2}+o((\Delta_N)^{m/2})\right),(C_3(\Delta_N))^m(\Delta_N)^{m/2-1}\right\}T}$$

where $K(m,\Delta_N)$ and $f(m,\Delta_N)$ are defined as in Proposition 2.4.1 and M_0 is a positive constant.

The proof of this corollary is given in Appendix A.1. Based on Corollary 2.4.1, the following Schwartz information based criterion function with penalty term is proposed:

$$\mathrm{IC}(p)=-2\log\mathcal{L}_N([0,T],\hat{\theta}_T(p))+(p+1)\log(N). \tag{2.4.6}$$

For short, let $\hat{\theta}(p)$ be $\hat{\theta}_T$ with dimension p, by the Riemann sum approximation of (2.3.1), $\mathrm{IC}(p)=-2\left(\log\mathcal{L}_N([0,T],\hat{\theta}_T(p))\right)+(p+1)\log(N)$. This gives

$\mathrm{IC}(p)=-2\left(\frac{1}{2\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left(-(Y_i-Z_i(p)\hat{\theta}(p))^2+(Y_i)^2\right)\right)+(p+1)\log(N)$. $Y_i =$

$Z_i(p)\theta(p) + \epsilon_i$ implies that

$$\mathrm{IC}(p) = -2\left(\frac{1}{2\sigma^2}\left(\sum_{i\in\mathbb{N}[0,T]}\frac{(Y_i)^2}{t_{i+1}-t_i} - \sum_{i\in\mathbb{N}[0,T]}\frac{\left(Z_i(p)\theta(p)+\epsilon_i - Z_i(p)\hat{\theta}(p)\right)^2}{t_{i+1}-t_i}\right)\right)+(p+1)\log(N).$$

Suppose that p_0 is the exact value of the number of the base functions in the SDE (2.2.1). To compare the value of $\mathrm{IC}(p_0)$ and $\mathrm{IC}(p)$ for $p \neq p_0$, first, the following six useful propositions are proven.

Proposition 2.4.2. *If Assumption 2.1-2.4 hold, then* $\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i)$ *is a positive definite matrix for* $T \geqslant 1$. *Further,* $\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} \Sigma$.

The proof of this proposition is given in Appendix A.1. Let $\gamma_1(T)$ be the smallest eigenvalue of matrix $\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i)$ and let γ_1 be the smallest eigenvalue of the matrix Σ. The following corollary gives the convergence of the eigenvalue $\gamma_1(T)$.

Corollary 2.4.2. *If Assumption 2.1-2.4 hold, then,* $\gamma_1(T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \gamma_1$.

Proof. The proof follows directly from Lemma A.5 in Appendix A.2. □

For the situation of unknown p, first, some notations related to p are given. Let p_* be a positive integer, which is less than p_0 and let p^* be a positive integer, which is greater than p_0. Let $\hat{\theta}(p_*+) = (\hat{\mu}_1, \hat{\mu}_2, \cdots, \hat{\mu}_{p_*}, 0_{p_*+1}, \cdots, 0_{p_0}, \hat{\alpha})^\top$. The following Lemma is derived, which is useful in proving that the log-likelihood-based information criterion $\mathrm{IC}(p)$ reaches its minimum value at the exact dimension p_0. Proposition 2.3.5 gives the limiting distribution of the vector random process $\frac{1}{\sqrt{T}}W_{[0,T]}(p)$. In the following proposition, it is proven that, for some $0 < a^* < \boldsymbol{a}/2$, $\left(\frac{1}{\sqrt{T}}\|W_{[0,T]}(p)\|\right) / \left(\log^{a^*}(T)\right)$ is bounded in probability. This is another result which plays an important role in proving that the log-likelihood-based information criterion $\mathrm{IC}(p)$ reaches its minimum value at the exact dimension p_0.

Proposition 2.4.3. *If Assumption 2.1-2.3, and Assumption 2.4 hold, then, for some* $0 < a^* < \boldsymbol{a}/2$*, for each* $p \geqslant 1$*,*

$$\frac{1}{\sqrt{T}}\|W_{[0,T]}(p)\| = O_p(\log^{a^*}(T)) \text{ and } \frac{1}{T}\|W_{[0,T]}(p)\|^2 = O_p(\log^{2a^*}T). \tag{2.4.7}$$

The proof of this proposition is provided in Appendix A.3. The following proposition shows that, for some $0 < a^* < \boldsymbol{a}/2$, $\left(\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,T]}\varepsilon_i Z_i(p)/(t_{i+1}-t_i)\right\|\right)\Big/\left(\log^{a^*}(T)\right)$ is bounded in probability, which play the same role as Proposition 2.4.3.

Proposition 2.4.4. *If Assumption 2.1-2.3, and Assumption 2.4 hold, then, for some* $0 < a^* < \boldsymbol{a}/2$*, for each* $p \geqslant 1$

$$\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,T]}\varepsilon_i Z_i(p)/(t_{i+1}-t_i)\right\| = O_p(\log^{a^*}(T)). \tag{2.4.8}$$

The proof of this proposition is given in Appendix A.1. Based on the previous propositions and lemmas, below, the main result of this subsection is derived. The following proposition shows that the Schwartz information criterion reaches its minimum value at the exact parameter dimension p_0 and it shows that the estimator obtained from the criterion is consistent. Let $\hat{p} = \arg\min_{p\in\mathbb{N}} \mathrm{IC}(p)$.

Proposition 2.4.5. *If Assumption 2.1-2.4 hold, then,* $\lim_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(p_0) < \mathrm{IC}(p)) = 1$, for $1 \leqslant p \neq p_0$ *and* $\hat{p} - p_0 \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0$.

The proof of this proposition is given in Appendix A.1. Next proposition shows that the proposed method improves the goodness-of-fit. For the partition $0 = t_0 < t_1 < t_2 < \cdots < t_{N-1} < t_N = T$, let $\varphi_{\hat{p}}(t_i) = \left(\varphi_1(t_i), \varphi_2(t_i), \cdots, \varphi_{\hat{p}}(t_i)\right)$ and $\varphi_{p_0}(t_i) = \left(\varphi_1(t_i), \varphi_2(t_i), \cdots, \varphi_{p_0}(t_i)\right)$. For $\hat{p} > p_0$, an auxiliary vector $\varphi_{p_0+}(t_i) = (\varphi_1(t_i), \varphi_2(t_i), \cdots, \varphi_{p_0}(t_i), 0, \cdots, 0_{\hat{p}})$ is constructed. For $\hat{p} < p_0$, the auxiliary vector $\varphi_{\hat{p}+}(t_i) = (\varphi_1(t_i), \varphi_2(t_i),$

$\cdots, \varphi_{\hat{p}}(t_i), 0, \cdots, 0_{p_0})$ is introduced. The auxiliary vectors are constructed without changing their norms and directions, they are just put into another vector space with different dimensions. From Proposition A.1-Corollary A.2, the established result shows that the proposed method improves the goodness-of-fit. To this end, let SSE(p) be the sum of squared error of the SIC (2.4.6) with dimension parameter p.

Proposition 2.4.6. *If Assumption 2.1-2.4 hold, then,* $\forall\ p \neq p_0$,

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}\left(\frac{1}{N}(\mathrm{SSE}(p) - \mathrm{SSE}(\hat{p})) \geqslant 0\right) = 1.$$

The proof of this proposition is given in Appendix A.1. In other words, Proposition 2.4.6 shows, as T is large and Δ_N is arbitrary small, that SSE(p) achieves its minimum value at $\hat{p}$ which indicates that the goodness-of-fit is the highest when the value of p is taken as $\hat{p}$.

2.4.2 The case of a possible unknown change-point

This subsection considers the case where both the location of change point t^* and the number of base functions p are unknown. To determine the estimators of p and t^*, the *SIC information criterion* which is used in Section 2.4.1 is slightly modified. As in the case of no change-point, let

$$Y_i = (X_1(t_{i+1}) - X_1(t_i))/X_1(t_i)\mathbb{I}_{t_i\in[0,sT]} + (X_2(t_{i+1}) - X_2(t_i))/X_2(t_i)\mathbb{I}_{t_i\in(st,T]}$$

for $i \in \mathbb{N}[0, T]$, and let

$$Z_i(p) = (Z_{1i}^\top(p), Z_{2i}^\top(p))^\top \quad \text{with,} \quad Z_{ki}'(p) = (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_p(t_i), -\ln X_k(t_i)(t_{i+1}-t_i), \tag{2.4.9}$$

for $k = 1, 2$, $i \in \mathbb{N}[0, T]$. From the discretized version of the process in (2.3.12),

$$Y_i = Z_i(p)\boldsymbol{\theta}(p) + \epsilon_i, \quad i \in \mathbb{N}[0, T], \tag{2.4.10}$$

with $\boldsymbol{\theta}(p) = \boldsymbol{\theta}$ with dimension $2(p+1)$, and ϵ_i the error term given by (2.4.4). Obviously, $\epsilon_1, \epsilon_2, \epsilon_3, \ldots$ are independent with $\epsilon_i \sim \mathcal{N}(0, \sigma^2 (t_{i+1} - t_i))$, $i \in \mathbb{N}[0, T]$. More precisely, in this case, the most appropriate model is the one that minimises the log-likelihood-based information criterion given by

$$\mathrm{IC}(c, p) = -2\log\mathcal{L}_N(\hat{t}^*, [0, T], \hat{\theta}_T(s)) + (c + 1)h(p)\Phi(T) \tag{2.4.11}$$

where, p is the dimension of $\varphi(t)$, $h(p)$ is the function of the parameter p, $\Phi(T)$ is a non-decreasing function of T, $c = 0$ or 1 is the number of change-point in the process (2.3.12), the function $\log\mathcal{L}_N(\hat{t}^*, [0, T], \hat{\theta}_T)$ is the Riemann sum approximation of the log-likelihood function $\log\mathcal{L}(\hat{t}^*, [0, T], \hat{\theta}_T)$ which is defined in (2.3.17), and $\hat{t}^*$ is given by

$$\hat{t}^* = \arg\max_{t^*}(\log\mathcal{L}_N(t^*, [0, T], \hat{\theta}_T(s))), \tag{2.4.12}$$

for a fixed value of p. Let $c^0 = 0$, or 1, represents the exact number of change-point and p_0 is the true value of the parameter p. The primary result of this subsection is that SIC (2.4.11) reaches its minimum at the value $p = p_0, c = c^0$. Let $c^0 = 0$, or 1, represent the exact number of change-point and p_0 is the true value of the parameter p. Note that, for a known p_0, the restriction (2.3.14) and $p = p_0$ lead to the following hypothesis testing problem:

$$H_0 : \tilde{M}(c, p)\boldsymbol{\theta} = \tilde{r} \quad \text{and} \quad p = p_0 \quad \text{versus} \quad H_1 : \tilde{M}(c, p)\boldsymbol{\theta} \neq \tilde{r} \;\; or \;\; p \neq p_0 \tag{2.4.13}$$

$\tilde{M}(c, p)$ is $q \times (c + 1)(p + 1)$ full rank matrix with $q < (c + 1)(p + 1)$, r is a known q-column vector. The matrix $\tilde{M}(c, p)$ has known components but unknown number of column. Let s^0 be the exact location of the change-point, and $\hat{s}$ is an estimator of s^0. Before estimating the dimension p, first, the following proposition is proven, which shows that $\hat{s}$ is a consistent estimator of s^0.Furthermore, let

$$\mathcal{Z}_{1i}(p) = \begin{cases} (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_p(t_i), -\ln X_1(t_i)(t_{i+1} - t_i), & \text{if } i \in \mathbb{N}[0, \hat{s}T] \text{ and } \hat{s} < s, \\ (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_p(t_i), -(\ln X_1(t_i)\mathbb{I}_{\{i\in\mathbb{N}[0,sT]\}} & \\ + \ln X_2(t_i)\mathbb{I}_{\{i\in\mathbb{N}[sT,\hat{s}T]\}})(t_{i+1} - t_i), & \text{if } i \in \mathbb{N}[0, \hat{s}T] \text{ and } \hat{s} > s, \end{cases}$$

and

$$\mathcal{Z}_{2i}(p) = \begin{cases} (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_p(t_i), -\ln X_2(t_i)(t_{i+1} - t_i), & \text{if } i \in \mathbb{N}[\hat{s}T, T] \text{ and } \hat{s} > s, \\ (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_p(t_i), -(\ln X_1(t_i)\mathbb{I}_{\{i\in\mathbb{N}[\hat{s}T,sT]\}} & \\ + \ln X_2(t_i)\mathbb{I}_{\{i\in\mathbb{N}[sT,T]\}})(t_{i+1} - t_i), & \text{if } i \in \mathbb{N}[\hat{s}T, T]] \text{ and } \hat{s} < s. \end{cases}$$

Let $\mathcal{Z}_i(p) = (\mathcal{Z}_{1i}^\top(p), \mathcal{Z}_{2i}^\top(p))^\top$.

Proposition 2.4.7. *Suppose that Assumption 2.1-2.3, and Assumption 2.4 hold and the shift in the drift parameters is of fixed non zero magnitude independent of T. Then,* $\hat{s} - s^0 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$.

The proof is given in proposition A.1. In Appendix A.3, Lemma A.6-Lemma A.7 are established, which show that $\frac{1}{\sqrt{T}} W_{[0,\hat{s}T]}(p) - \frac{1}{\sqrt{T}} W_{[0,sT]}(p) \xrightarrow[T\to\infty]{\mathrm{P}} 0$. Together with Proposition A.18 in Appendix A.3 imply that $\hat{\theta}_T(\hat{s}) \xrightarrow[T\to\infty]{a.s.} \boldsymbol{\theta}$. Further, this result with Proposition 2.3.18 gives

$$\hat{\theta}_T(\hat{s}) - \hat{\theta}_T(s) \xrightarrow[T\to\infty]{a.s.} 0. \tag{2.4.14}$$

The following proposition is a remarkable result, which shows that the improved *Schwartz information criterion* reaches its minimum value at the point (c^0, p_0). Let $(\hat{p},\ \hat{c}) = \arg \min_{c\in\{0,1\}, p\in\mathbb{N}_+} \mathrm{IC}(c, p)$.

Proposition 2.4.8. *If Assumption 2.1-2.3, and Assumption 2.4 hold, then,*

$$\lim_{T\to\infty} \mathrm{P}(\mathrm{IC}(c^0, p_0) > \mathrm{IC}(c, p)) = 0, \text{for any} c \neq c^0 \text{or} p \neq p_0. \tag{2.4.15}$$

Further,

$$\hat{p} - p_0 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0, \quad \hat{c} - c^0 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0.$$

The proof of this result is given in Appendix A.1. To conclude this subsection, the hypothesis testing problem in (2.4.13) is tackled. The following set of local alternatives restrictions is under consideration:

$$H_{\tilde{a},T} : \tilde{M}(c,p)\boldsymbol{\theta} - \tilde{r} = \frac{\tilde{r}_0}{\sqrt{T}}, T > 0, \tag{2.4.16}$$

where $\tilde{M}(c,p)$ is a $q \times (c+1)(p+1)$ full rank matrix with known elements and $q < (c+1)(p+1)$, r is a known q column vector. From the testing problem (2.4.13) and from the local alternative restrictions (2.4.16), to study the asymptotic normality of $\tilde{\theta}_T(s)$, define

$$\tilde{\Sigma}_c(p) = \begin{cases} \Sigma(p), & if\ c = 0 \\ \tilde{\Sigma}(p), & if\ c = 1, \end{cases} \qquad Q_c(s,T,p) = \begin{cases} Q_{[0,T]}(p), & if\ c = 0 \\ Q(s,T,p), & if\ c = 1, \end{cases}$$

where $\Sigma(p)$ is, as defined in (2.3.6), a size $(p+1)\times(p+1)$-matrix and $\tilde{\Sigma}(p)$ is, as defined in (2.3.20), $2(p+1)\times 2(p+1)$-matrix. Let

$$\Delta = \frac{1}{\sigma^2}\tilde{r}_0^\top \left(\tilde{M}(c,p)\tilde{\Sigma}_c^{-1}(p)\tilde{M}^\top(c,p)\right)^{-1}\tilde{r}_0, \tag{2.4.17}$$

$$\begin{aligned} \Gamma(c,p) &= \frac{1}{\sigma^2}\tilde{M}^\top(c,p)\left(\tilde{M}(c,p)\tilde{\Sigma}_c^{-1}(p)\tilde{M}^\top(c,p)\right)^{-1}\tilde{M}(c,p), \\ \hat{\Gamma}(c,p) &= \frac{1}{\hat{\sigma}^2}\tilde{M}^\top(c,p)\left(\tilde{M}(c,p)TQ_c^{-1}(s,T,p)\tilde{M}^\top(c,p)\right)^{-1}\tilde{M}(c,p). \end{aligned} \tag{2.4.18}$$

Let $\tilde{\theta}_T(\hat{s},\hat{c},\hat{p})$, $\hat{\theta}_T(\hat{s},\hat{c},\hat{p})$ be the UMLE and RMLE of $\boldsymbol{\theta}$ based on the estimation of s, c and p. Let $\tilde{G}_{[0,T]}(\hat{s},\hat{c},\hat{p}) = Q_{\hat{c}}^{-1}(\hat{s},T,\hat{p})\tilde{M}^\top(\hat{c},\hat{p})\left(\tilde{M}(\hat{c},\hat{p})Q_{\hat{c}}^{-1}(\hat{s},T,\hat{p})\tilde{M}^\top(\hat{c},\hat{p})\right)^{-1}$. As in Proposition 2.3.15,

$$\tilde{\theta}_T(\hat{s},\hat{c},\hat{p}) = \hat{\theta}_T(\hat{s},\hat{c},\hat{p}) - \tilde{G}_{[0,T]}(\hat{s},\hat{c},\hat{p})\left(\tilde{M}(\hat{c},\hat{p})\hat{\theta}_T(\hat{s},\hat{c},\hat{p}) - \tilde{r}\right). \tag{2.4.19}$$

Let $\tilde{G}^*(c^0,p_0) = \tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}$. Before constructing the statistics of the testing problem (2.4.13), the following proposition is presented.

To introduce some notations, let $\tilde{\varsigma} \sim \mathcal{N}_{(c^0+1)(p_0+1)}\left(\tilde{G}^*(c^0, p_0)\tilde{r}_0, \sigma^2\tilde{G}^*(c^0, p_0)\tilde{M}(c^0, p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)$, $\tilde{\varsigma}_0 \sim \mathcal{N}_{(c^0+1)(p_0+1)}\left(0, \sigma^2\tilde{G}^*(c^0, p_0)\tilde{M}(c^0, p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)$, $\psi(s, c^0, p_0) = \tilde{\varsigma}^\top\Gamma(c^0, p_0)\tilde{\varsigma}$, $\psi_0(s, c^0, p_0) = \tilde{\varsigma}_0^\top\Gamma(c^0, p_0)\tilde{\varsigma}_0$, and let

$$\tilde{\varsigma}_T(\hat{s}, \hat{c}, \hat{p}) = \sqrt{T}(\hat{\theta}(\hat{s}, \hat{c}, \hat{p}) - \tilde{\theta}(\hat{s}, \hat{c}, \hat{p})), \quad \tilde{\psi}_T(\hat{s}, \hat{c}, \hat{p}) = \tilde{\varsigma}_T^\top(\hat{s}, \hat{c}, \hat{p})\hat{\Gamma}(\hat{c}, \hat{p})\tilde{\varsigma}_T(\hat{s}, \hat{c}, \hat{p}). \tag{2.4.20}$$

The following presented proposition is useful in solving the testing problem in (2.4.13).

Proposition 2.4.9. *If Assumption 2.1-2.3, and Assumption 2.4 hold, then, if* $\tilde{r}_0 \neq 0$, $\tilde{\psi}_T(\hat{s}, \hat{c}, \hat{p}) \xrightarrow[T\to\infty]{\mathrm{D}} \psi(s, c^0, p_0) \sim \chi_q^2(\Delta)$. *If* $\tilde{r}_0 = 0$, *then,* $\tilde{\psi}_T(\hat{s}, \hat{c}, \hat{p}) \xrightarrow[T\to\infty]{\mathrm{D}} \psi_0(s, c^0, p_0) \sim \chi_q^2$.

The proof is given in Appendix A.1. In this case, the null hypothesis in (2.4.13) is tested by using the rejection region $\tilde{\psi}_T(\hat{c}, \hat{p}) > \chi_{\alpha;q}^2$ for a given α, i.e. the suggested test is

$$\tilde{\kappa}_T(\hat{s}, \hat{c}, \hat{p}) = \mathbb{I}_{\{\tilde{\psi}_T(\hat{c}, \hat{p}) > \chi_{\alpha;q}^2\}} \tag{2.4.21}$$

From Proposition 2.3.8, below, the asymptotic power of the proposed test is established.

Proposition 2.4.10. *If Assumption 2.1-2.3, and Assumption 2.4 hold, then, the asymptotic power function of the test in (2.4.21) is given by* $\Pi(\Delta) = \mathrm{P}\left(\chi_q^2(\Delta) \geqslant \chi_{\alpha;q}^2\right)$.

Proof. The proof follows directly from Proposition A.11 and Proposition 2.4.9. □

2.5 Shrinkage estimators and comparison between estimators

This section presents shrinkage estimators (SEs) which combine in an optimal way the UMLE and the RMLE. As frequently noticed in constrained inference, if the restriction is not correct, the UMLE performs better than the RMLE while if the restriction holds, the RMLE dominates the UMLE. However, more often than not, it is not possible to be totally sure about the validity of the restriction. Thus, it is important to derive a statistical method which is robust with respect to the restriction. The SEs have the advantage of preserving a very good performance regardless of the validity of the restriction. Nevertheless, since the dimensions of the UMLE and the RMLE are random, the derivation of shrinkage estimators as well as their relative efficiency do not follow from the results in classical literature. In particular, the following class of shrinkage type estimators are under consideration

$$\hat{\theta}_T^s(\hat{s},\hat{c},\hat{p}) = \tilde{\theta}_T(\hat{s},\hat{c},\hat{p}) + \gamma\left(T\left\|\hat{\theta}_T(\hat{s},\hat{c},\hat{p}) - \tilde{\theta}_T(\hat{s},\hat{c},\hat{p})\right\|_{\hat{\Gamma}}^2\right)\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}) - \tilde{\theta}_T(\hat{s},\hat{c},\hat{p})\right) \tag{2.5.1}$$

where $\|x\|_A^2 = \text{trace}\,(x^\top A x)$ given x is a column vector, γ is continuous real-valued function on $(0,+\infty)$ and $\hat{\Gamma}$ is defined in (2.4.18). It is obvious that if $\gamma(x) = 0$, $\hat{\theta}_T^s(\hat{s},\hat{c},\hat{p}) = \hat{\theta}_T^{s+}(\hat{s},\hat{c},\hat{p}) = \tilde{\theta}_T(\hat{s},\hat{c},\hat{p})$, if $\gamma(x) = 1$, $\hat{\theta}_T^s(\hat{s},\hat{c},\hat{p}) = \hat{\theta}_T^{s+}(\hat{s},\hat{c},\hat{p}) = \hat{\theta}_T(\hat{s},\hat{c},\hat{p})$. As an example, if $\gamma(x) = 1 - \frac{q-2}{x}$, with $3 \leqslant q = \text{rank}(\tilde{M}(\hat{c},\hat{p})) < (\hat{c}+1)(\hat{p}+1)$. The shrinkage estimators (SEs) is given as

$$\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}) = \tilde{\theta}_T(\hat{s},\hat{c},\hat{p}) + \left(1 - (q-2)\tilde{\psi}_T^{-1}\right)\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}) - \tilde{\theta}_T(\hat{s},\hat{c},\hat{p})\right), \tag{2.5.2}$$

where $\tilde{\psi}_T$ is defined in (2.4.20). To avoid an over-shrinking problem, by taking $\gamma(x) = [1 - \frac{q-2}{x}]^+$, $x > 0$, the positive-part shrinkage estimator (PSE) is given as

$$\hat{\theta}_T^{sh+}(\hat{s},\hat{c},\hat{p}) = \tilde{\theta}_T(\hat{s},\hat{c},\hat{p}) + \left[1 - (q-2)\tilde{\psi}_T^{-1}\right]^+\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}) - \tilde{\theta}_T(\hat{s},\hat{c},\hat{p})\right). \tag{2.5.3}$$

As commonly the case in large sample-point estimation, to study the relative efficiency of the proposed estimators, the criterion known as Asymptotic Distributional Risk (ADR) is used. For the convenience of the reader, first this concept is recalled. The Asymptotic Distributional Risk (ADR) of an estimator $\hat{\theta}_0(\hat{s},\hat{c},\hat{p})$ is defined as

$$\text{ADR}\left(\hat{\theta}_0(\hat{s},\hat{c},\hat{p}), \boldsymbol{\theta}; \Omega\right) = \mathbb{E}[(\varepsilon^\top \Omega \varepsilon)] \tag{2.5.4}$$

where ε is a random vector such that $\sqrt{T}(\hat{\theta}_0(\hat{s},\hat{c},\hat{p})-\boldsymbol{\theta})^\top \sqrt{T}(\hat{\theta}_0(\hat{s},\hat{c},\hat{p})-\boldsymbol{\theta})\mathbb{I}_{\{\hat{c}=c^0,\hat{p}=p_0\}} \xrightarrow[T\to\infty]{D}$ $\varepsilon^\top\varepsilon$, and Ω is a $(c^0+1)(p_0+1)\times(c^0+1)(p_0+1)$-positive symmetric semi-definite weighting matrix. It should be noticed that the concept of the ADR used here is slightly different to that used for example in Saleh [2006], Nkurunziza and Zhang [2018] and Nkurunziza [2012] among others. Indeed, in the quoted papers, the dimensions of the proposed estimators are nonrandom while in this chapter, the dimensions of the proposed estimators are random. Under the criterion in (2.5.4), the following proposition is derived, which shows that, near the restriction, the RMLE dominates the UMLE while the UMLE is better than the RMLE as one moves far away from the restriction. Let $\Lambda = \tilde{\Sigma}_{c^0}^{-1}(p_0) - \tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)$. Let $\lambda_{\min},\lambda_{\max}$ be the smallest and largest eigenvalues of the matrix $\left(\tilde{G}^{*\top}(c^0,p_0)\Gamma\tilde{G}^*(c^0,p_0)\right)^{-1}\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)$, respectively.

Proposition 2.5.1. *If Assumption 2.1-2.3, and Assumption 2.4 along with the set of local alternatives in (2.4.16) hold,*

1. *if* $\Delta \leqslant \sigma^2\mathrm{trace}(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\boldsymbol{\Sigma}_{c^0}^{-1}(p_0))/\lambda_{\max}$, *then,*
$$\mathrm{ADR}\left(\tilde{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant \mathrm{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right);$$
2. *if* $\Delta \geqslant \sigma^2\mathrm{trace}(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\boldsymbol{\Sigma}_{c^0}^{-1}(p_0))/\lambda_{\min}$, *then,*
$$\mathrm{ADR}\left(\tilde{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \geqslant \mathrm{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right).$$

The proof of this proposition is given in Appendix A.4. The following proposition indicates that the SEs dominate the UMLE. Further, it shows that PSE dominates the SEs.

Proposition 2.5.2. *If Assumption 2.1-2.3, and Assumption 2.4 along with the local al-*

ternative restriction (2.4.16) hold, then,

$$\text{ADR}\left(\hat{\theta}_T^{sh+}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right),$$

for all $\Delta \geqslant 0$, *provided* $2\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right) \geqslant (q+2)\lambda_{\max}$.

The proof of this proposition is given in Appendix A.4.

2.6 Simulation study and real dataset

In this section, Monte-Carlo simulation technique is used to evaluate the performance of the proposed method. The proposed method is also applied to a real data set. In particular, the gold spot historical price dataset is analized.

Several cases have been explored but, in order to save the space of this chapter, here only the results obtained from the case where there is 1 change-point with $p = 2$ is reported. The GEOU process is generatedby the same trigonometric orthogonal function system.

1. If $p = 1$, the basis is $\{1\}$.
2. If $p = 2$, the basis is $\{1,\ \sqrt{2}\cos(\omega t)\}$.
3. If p is odd and greater than 2, the basis is given as
$$\left\{1,\ \sqrt{2}\cos(\omega t),\ \sqrt{2}\sin(\omega t),\cdots,\ \sqrt{2}\cos\left(\frac{p-1}{2}\omega t\right),\ \sqrt{2}\sin\left(\frac{p-1}{2}\omega t\right)\right\}.$$
4. If p is even and greater than 2, the basis is given as
$$\left\{1,\ \sqrt{2}\cos(\omega t),\ \sqrt{2}\sin(\omega t),\cdots,\ \sqrt{2}\cos\left(\frac{p-2}{2}\omega t\right),\ \sqrt{2}\sin\left(\frac{p-2}{2}\omega t\right),\ \sqrt{2}\cos\left(\frac{p}{2}\omega t\right)\right\}.$$

where $\omega = 2\pi$.

2.6.1 Simulation with no change point case

Specifically, for the case where $p = 4$, set $\theta = (3, 1, 2, 1, 1)'$ and to specify the restriction, let $M = [0_{3\times1} \ \vdots \ I_3 \ \vdots \ 0_{3\times1}]$. Under the restriction, by using the proposed method, the point estimates of UMLE, the RMLE and the SEs are computed, as well as the standard error of each estimate. The obtained numerical results are reported in Table 2.1-Table 2.2 for the cases where $T = 35$ and $T = 50$. Overall, the simulation results show that as the time horizon T increases, the estimators are closer to the exact value of pre-assigned coefficients. Also, as the time horizon increases, the standard errors are getting smaller as T increases. Figure 2.1 gives the histogram of the estimators when

Table 2.1: Mean and standard deviation of estimators of drift parameters (T=35)

Parameters	μ_1	μ_2	μ_3	μ_4	α
UMLE	3.0112	0.9977	2.0009	0.9982	1.0038
	(0.2534)	(0.0303)	(0.0239)	(0.0172)	(0.0858)
RMLE	2.9924	1.0000	2.0000	1.0000	0.9974
	(0.1199)	(0.3945e-16)	(0.6365e-16)	(0.0000)	(0.0403)
SEs	2.9887	1.0008	1.9999	0.9994	0.9962
	(0.2614)	(0.0324)	(0.0222)	(0.0164)	(0.0884)
PSEs	3.0004	0.9991	2.0005	0.9993	1.0001
	(0.0163)	(0.0152)	(0.0124)	(0.0087)	(0.0561)

Table 2.2: Mean and standard deviation of estimators of drift parameters (T=50)

Parameters	μ_1	μ_2	μ_3	μ_4	α
UMLE	2.9963	0.9999	1.9999	0.9998	0.9987
	(0.2344)	(0.0265)	(0.0208)	(0.0144)	(0.0793)
RMLE	2.9935	1.0000	2.0000	1.0000	0.9978
	(0.1080)	(0.4715e-16)	(0.6365e-16)	(0.0000)	(0.0366)
SEs	3.0062	0.9999	2.0031	1.0009	1.0021
	(0.2369)	(0.0285)	(0.0217)	(0.0148)	(0.0800)
PSEs	2.9991	1.0003	2.0016	1.0007	0.9997
	(0.1499)	(0.0135)	(0.0106)	(0.0075)	(0.0507)

$T = 50$. The portray given by Figure 2.1 is consistent with the result given by Proposition 3.3.11. Indeed, the histograms seem quite symmetric concerning the pre-assigned values. Kolmogorov–Smirnov test is performed on the UMLE at $T = 50$, which corroborates the fact that UMLE is asymptotically normal. The data under the alternative

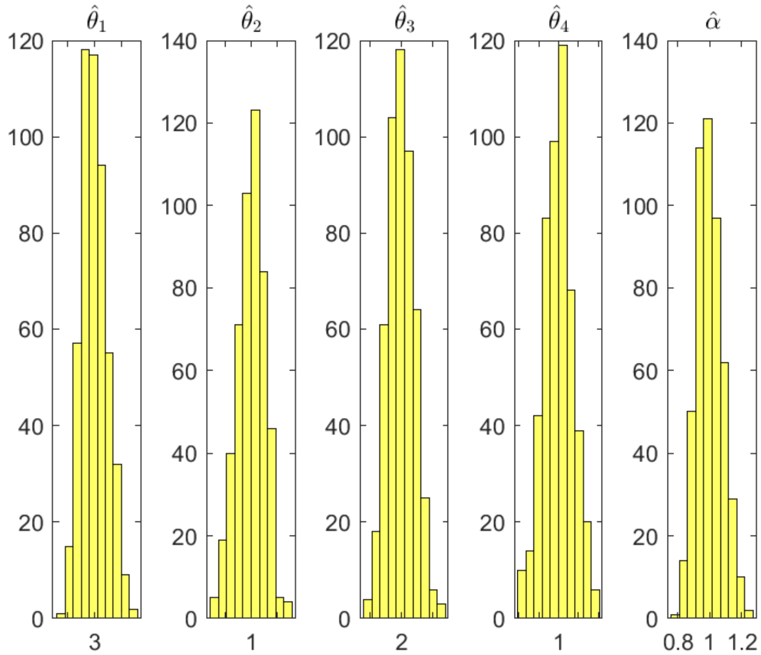

Figure 2.1: The histogram of estimators of GEOU model with no change point at $T = 50$

hypothesis is also generated. To this end, set $r_0 = 0.12kr$, $k = 1, 2, 3, 4, 5, 6$. Compute Δ as $\Delta = \frac{1}{\sigma^2} r_0^\top (M\Sigma^{-1}M^\top)^{-1} r_0$ and the Relative Mean Square Error (RMSE). For the cases where $T = 35$ and $T = 50$, the variation of the RMSE versus the non-centrality parameter is given by Figure 2.2 and Figure 2.3. These figures give a portray which confirms the theoretical result given in Proposition 3.5.3. Indeed, for each time horizon T, the plots of the RMSE versus the non-centrality parameter show that, near the null hypothesis, the RMLE has the best performance among all the four types of proposed estimators. However, as one moves far away from the null hypothesis, SEs dominate the RMLE. Further, the numerical findings confirm that the SEs are better than UMLE. The results for the case where $p = 4$, with the time horizon $T = 35$, $T = 50$, and $T = 50$ are reported. The obtained point estimates and their standard errors are reported in Table 2.1 and Table 2.2 and the variation of the RMSE versus the non-centrality parameter is given by Figure 2.2 and Figure 2.3.

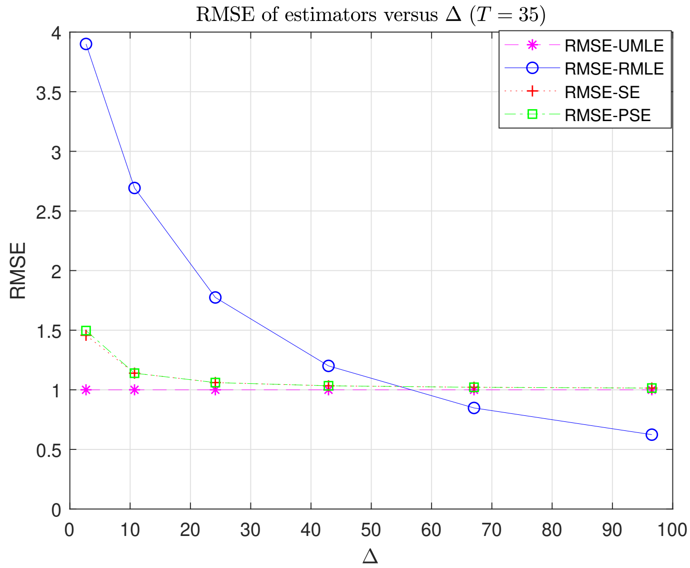

Figure 2.2: The RMSE of the estimator versus Δ ($T = 35$)

The proposed method is also applied to estimate the parameter dimension p. From 500 iterations, te the cumulative frequency (CF) and the relative frequency (RF) are computed, which are defined as $\mathrm{CF} = \sum_{i=1}^{500} \mathbb{I}_{\{\hat{p}_i = p\}}$, $\mathrm{RF} = \frac{1}{500} \sum_{i=1}^{500} \mathbb{I}_{\{\hat{p}_i = p\}} \times 100\%$, respectively. The obtained results, for the cases where $T = 20, 35, 50$ and $T = 80$, are shown in Table 2.3. To highlight the performance of the proposed test, see the reports in Figure 2.8-

Table 2.3: Cumulative frequency (CF) and the relative frequency (RF) of $\hat{p}$

Time	$T = 20$	$T = 35$	$T = 50$	$T = 80$
CF	499	500	500	500
RF	99.80%	100%	100%	100%

Figure 2.10 the variation of the empirical power versus the noncentrality parameter Δ under different significant level with $T = 80$. Let $M = [0_{2\times 1} \vdots I_2]$, when $p = 2$; Let $M = [0_{2\times 1} \vdots I_2 \vdots 0_{2\times 1}]$, when $p = 3$; Let $M = [0_{4\times 1} \vdots I_4 \vdots 0_{4\times 1}]$, when $p = 5$ and Let $M = [0_{5\times 1} \vdots I_5 \vdots 0_{5\times 1}]$, when $p = 6$. Δ was calculated with $r_0 = 0.04kr$, $k = 1, 2, 3, 4, 5, 6$.

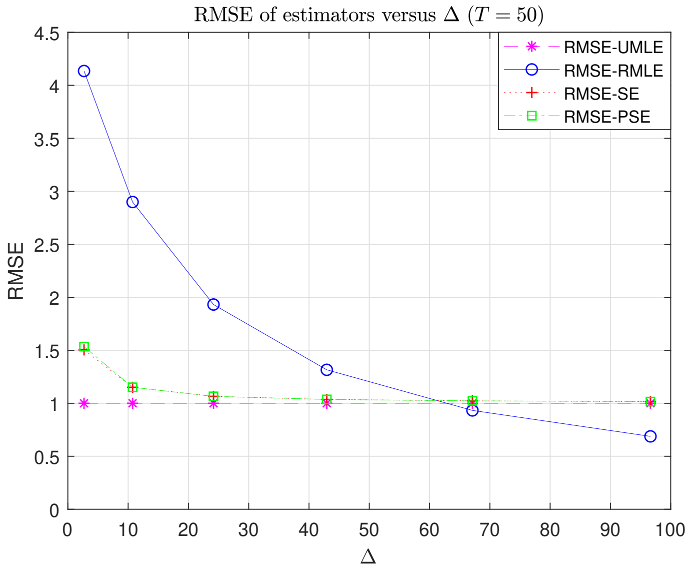

Figure 2.3: The RMSE of the estimator versus Δ ($T = 50$)

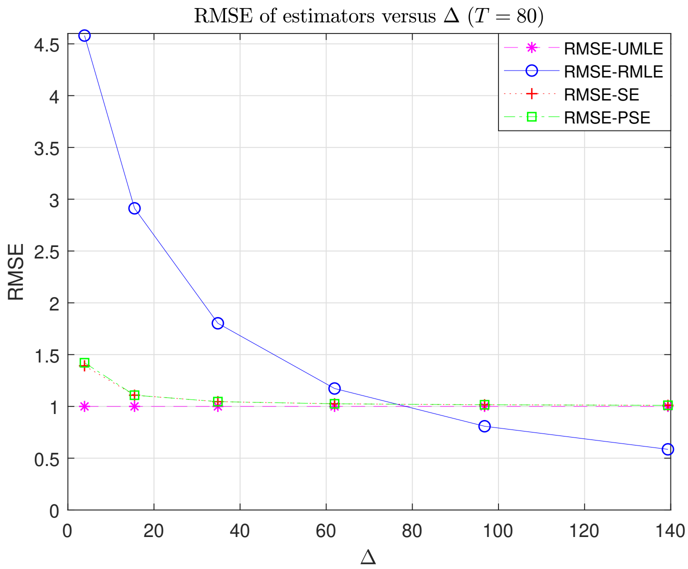

Figure 2.4: The RMSE of the estimator versus Δ ($T = 80$)

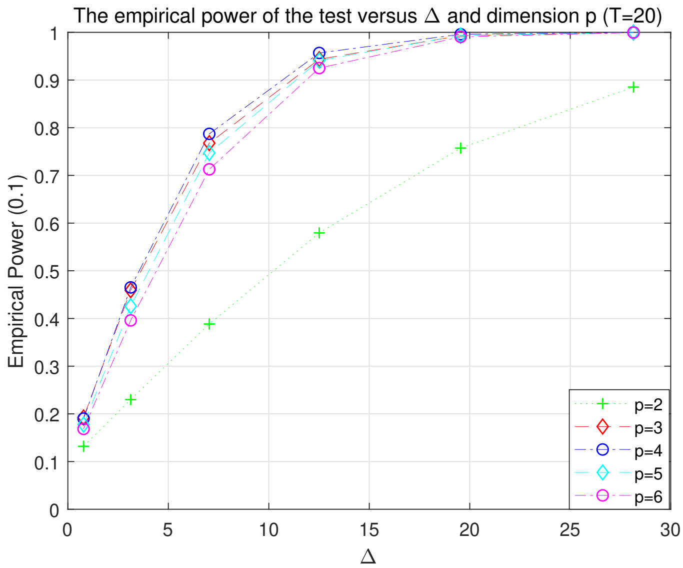

Figure 2.5: The empirical power of the test with different Δ and p ($T = 20, \alpha = 0.1$)

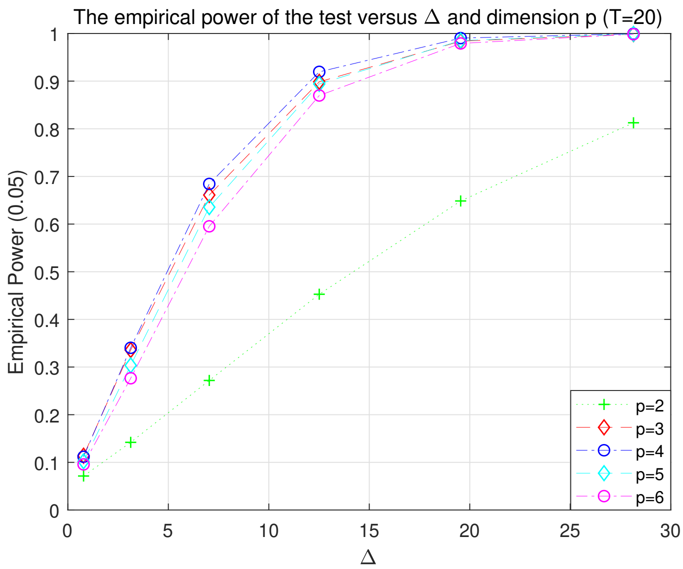

Figure 2.6: The empirical power of the test with different Δ and p ($T = 20, \alpha = 0.05$)

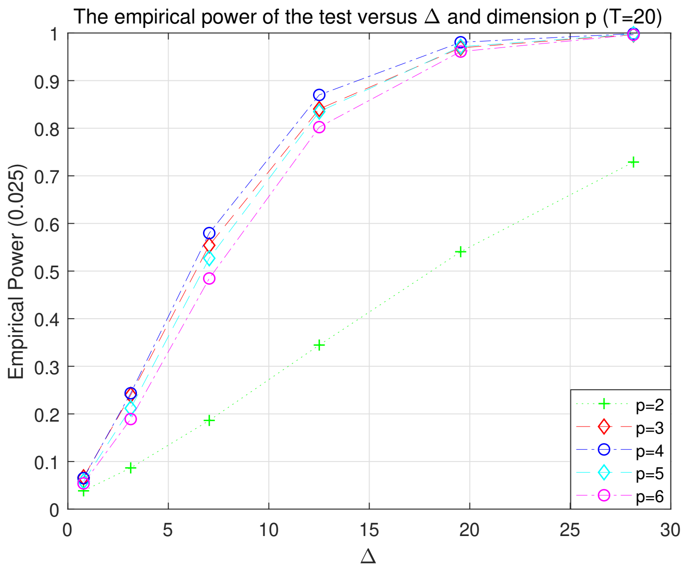

Figure 2.7: The empirical power of the test with different Δ and p ($T = 20, \alpha = 0.025$)

Figures 2.8- 2.10 indicate that the power of the test is the highest when p takes its exact value, i.e. if $p = 4$.

2.6.2 Simulation with one unknown change point

In this subsection, the simulation results are presented. Several cases have been explored but, in order to save the space of this chapter, here the results obtained from the case where there is 1 change-point with $p = 2$ is reported. In particular, for $p = 2$, the GEOU process is given by:

$$dX(t) = \sum_{i=1}^{2} \left(\mu_{i1} + \mu_{i2}\sqrt{2}\cos(2\pi t) - \alpha_i \ln X(t)\right) X(t)\mathbb{I}_{\{t_{i-1}^* \leqslant t \leqslant t_i^*\}}dt + \sigma X(t)dW_t. \quad (2.6.1)$$

where $t_0^* = 0$ and $t_2^* = T$ and t_1^* is the given change-point. To carry out the simulations, the pre-assigned value is $t_1^* = 0.5T$. The pre-assigned values for the drift parameter is $\boldsymbol{\theta} = (0.5, 0.5, 1, 2, 1, 2)^\top$. Let $\tilde{M} = \begin{bmatrix} -I_3, & I_3 \end{bmatrix}$. Under the restriction $\tilde{M}\boldsymbol{\theta} = 0$, the UMLE, the RMLE, the SEs, and their standard deviation within parentheses are reported in Table 2.4-Table 2.7. The results in Table 2.4-Table 2.7 show that, the time horizon T

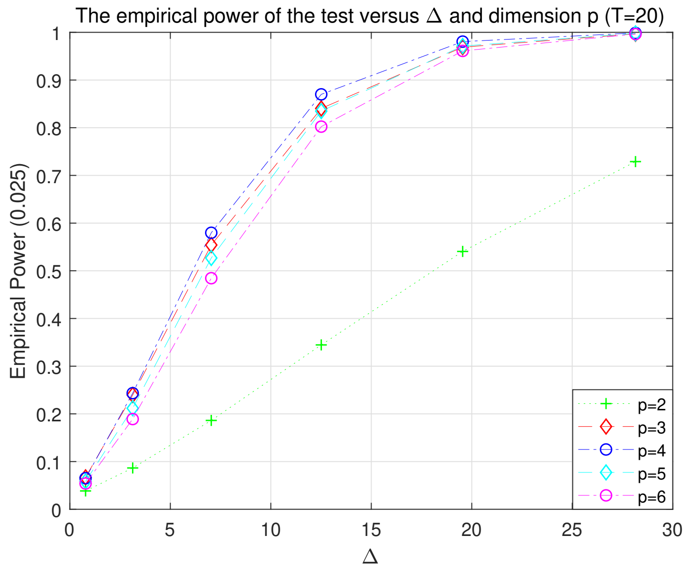

Figure 2.8: The empirical power of the test ($T = 80, \alpha = 0.025$)

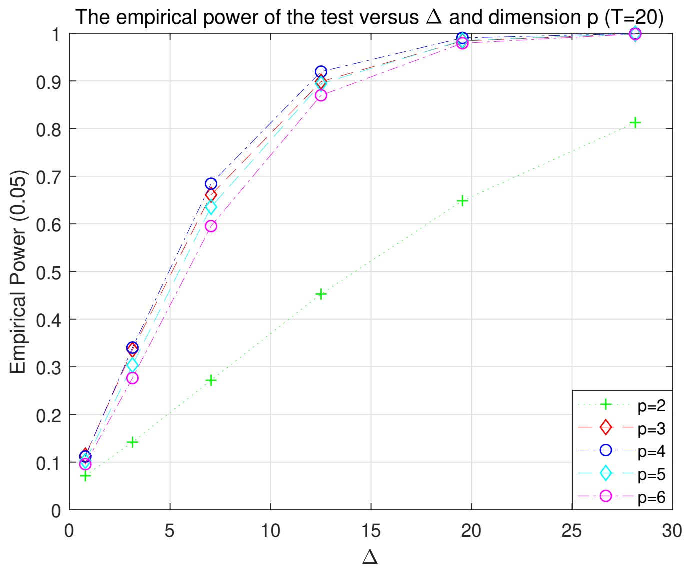

Figure 2.9: The empirical power of the test ($T = 80, \alpha = 0.05$)

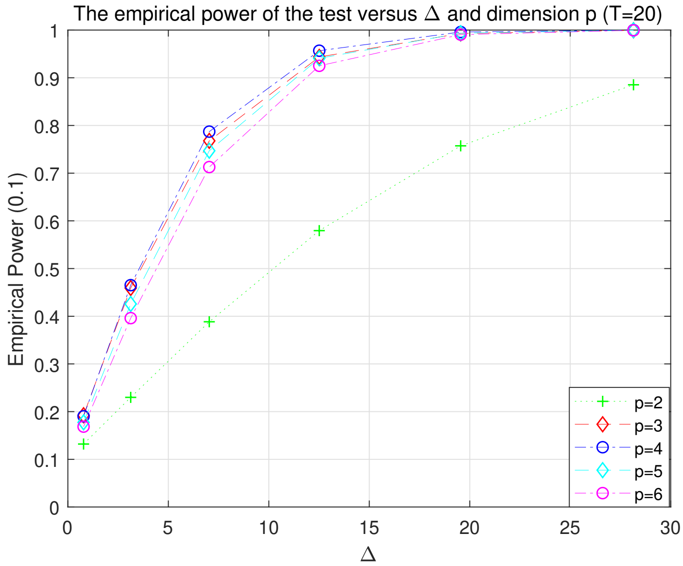

Figure 2.10: The empirical power of the test ($T = 80, \alpha = 0.1$)

increases, the estimators are closer to the exact value of pre-assigned coefficients. This can also been seen from the fact that, the standard errors are getting smaller as T increases. To show the advantage of the RMLE while the restriction, $r = \tilde{M}\boldsymbol{\theta}$, is correct, in Table 2.8 the *Relative Mean Squared Error* (RMSE) under this null hypothesis is reported. The numerical results given in Table 2.8 show that the RMLE has the best performance among all the four types of proposed estimators. This confirms the theoretical conclusion given in Proposition 3.5.3. Further, Table 3.5 shows that the SEs are better than UMLE, which is in agreement with the theoretical result given in Proposition 2.5.2. In Figure 2.11 the histogram of the estimates when $T = 80$ is reported. In order to save the space of this chapter, the histograms corresponding to the cases where $T = 20, 35, 50$ are not reported here but they have a similar visual portray.

Furthermore, by using the SIC (2.4.11), the dimension p and the number of change-point c are estimated. To estimate these parameters, the value of T needs to be larger than the ones used for the cases where these parameters are known. Let $\Delta_N = 1/100$ and from 500 iterations, compute the cumulative frequency (CF) and the relative frequency

Table 2.4: Mean and standard deviation of estimators of drift parameters (T=80)

Parameters	μ_{11}	μ_{12}	α_1	μ_{21}	μ_{22}	α_2
UMLE	0.5436	0.5009	1.0907	2.0983	1.0026	2.0961
	(0.1188)	(0.0337)	(0.2400)	(0.3049)	(0.0352)	(0.3065)
RMLE	0.2339	0.7068	0.2751	0.2339	0.7068	0.2751
	(0.0269)	(0.0226)	(0.0324)	(0.0269)	(0.0226)	(0.0324)
SEs	0.5400	0.5033	1.0813	2.0767	0.9992	2.0750
	(0.1176)	(0.0337)	(0.2376)	(0.3021)	(0.0353)	(0.3036)
PSE	0.5400	0.5033	1.0813	2.0767	0.9992	2.0750
	(0.1176)	(0.0337)	(0.2376)	(0.3021)	(0.0353)	(0.3036)

Table 2.5: Mean and standard deviation of estimators of drift parameters (T=50)

Parameters	μ_{11}	μ_{12}	α_1	μ_{21}	μ_{22}	α_2
UMLE	0.5508	0.5018	1.1231	2.1334	1.0060	2.1362
	(0.1343)	(0.0412)	(0.2876)	(0.3762)	(0.0455)	(0.3855)
RMLE	0.2474	0.7092	0.2864	0.2474	0.7092	0.2864
	(0.0336)	(0.0277)	(0.0415)	(0.0336)	(0.0277)	(0.0415)
SEs	0.5454	0.5055	1.1082	2.0998	1.0008	2.1033
	(0.1324)	(0.0412)	(0.2834)	(0.3712)	(0.0457)	(0.3803)
PSE	0.5454	0.5055	1.1082	2.0998	1.0008	2.1033
	(0.1324)	(0.0412)	(0.2834)	(0.3712)	(0.0457)	(0.3803)

Table 2.6: Mean and standard deviation of estimators of drift parameters (T=35)

Parameters	μ_{11}	μ_{12}	α_1	μ_{21}	μ_{22}	α_2
UMLE	0.5732	0.5011	1.1693	2.1756	1.0056	2.1793
	(0.1628)	(0.0480)	(0.3480)	(0.4401)	(0.0536)	(0.4453)
RMLE	0.2673	0.7042	0.3030	0.2673	0.7042	0.3030
	(0.0420)	(0.0320)	(0.0522)	(0.0420)	(0.0320)	(0.0522)
SEs	0.5656	0.5062	1.1477	2.1279	0.9982	2.1324
	(0.1595)	(0.0479)	(0.3406)	(0.4327)	(0.0538)	(0.4377)
PSE	0.5656	0.5062	1.1477	2.1279	0.9982	2.1324
	(0.1595)	(0.0479)	(0.3406)	(0.4327)	(0.0538)	(0.4377)

Table 2.7: Mean and standard deviation of estimators of drift parameters (T=20)

Parameters	μ_{11}	μ_{12}	α_1	μ_{21}	μ_{22}	α_2
UMLE	0.5886	0.5032	1.3213	2.2598	1.0088	2.2674
	(0.2182)	(0.0654)	(0.4708)	(0.5239)	(0.0683)	(0.5401)
RMLE	0.3134	0.7119	0.3425	0.3134	0.7119	0.3425
	(0.0665)	(0.0444)	(0.0844)	(0.0665)	(0.0444)	(0.0844)
SEs	0.6263	0.5112	1.2839	2.1840	0.9974	2.1924
	(0.2122)	(0.0651)	(0.4570)	(0.5107)	(0.0685)	(0.5259)
PSE	0.6263	0.5112	1.2839	2.1840	0.9974	2.1924
	(0.2122)	(0.0651)	(0.4570)	(0.5107)	(0.0685)	(0.5259)

Table 2.8: RMSE under $M\theta = (-1.5, -0.5, -1.0)^{\top}$

	$T = 20$	$T = 35$	$T = 50$	$T = 80$
RMSE-UMLE	1.0000	1.0000	1.0000	1.0000
RMSE-RMLE	11.4757	14.0767	13.5212	13.2343
RMSE-SEs	1.8814	1.9196	1.8649	1.8399
RMSE-PSE	2.0238	2.2341	2.1221	2.1852

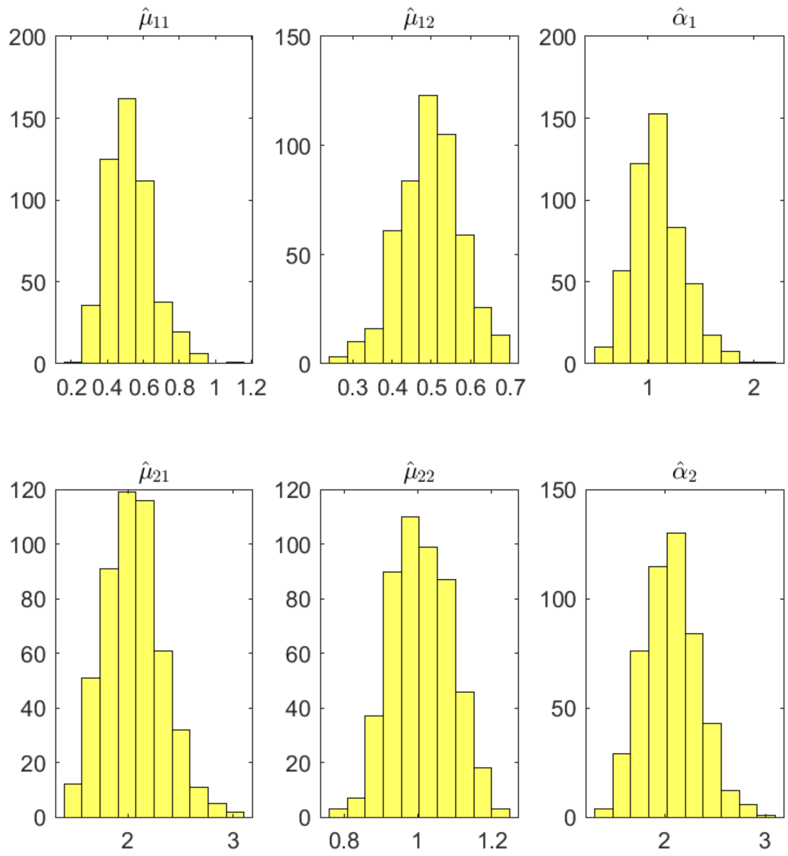

Figure 2.11: The histogram of estimators of GEOU model with one change point at $T = 80$

(RF) which are defined as CF $= \sum_{i=1}^{500} \mathbb{I}_{\{\hat{c}_i=c,\hat{p}_i=p\}}$, RF $= \frac{1}{500}\sum_{i=1}^{500} \mathbb{I}_{\{\hat{c}_i=c,\hat{p}_i=p\}} \times 100\%$. The obtained results, for the cases where $T = 20, 35$, and $T = 50$, are shown in the 3-D histograms (Figure 2.12-Figure 2.14) and Table 2.9. The results also show that RF is increasing to 1 as T tends to infinity.

Table 2.9: Cumulative frequency (CF) and the relative frequency (RF) of $(\hat{c}, \hat{p})$

Time	$T = 20$	$T = 35$	$T = 50$	$T = 80$
CF	432	487	493	494
RF	86.40%	97.4%	98.6%	98.8%

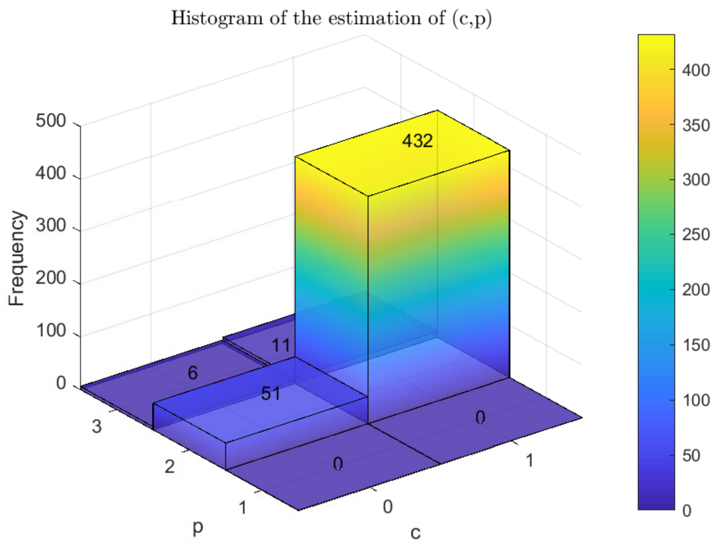

Figure 2.12: The histogram of the estimator of $(\hat{c}, \hat{p})$ $(T = 20)$

The behaviour of the empirical power of the proposed test versus different dimension p is analyzed. For $p = 1$, let $\tilde{M} = \begin{bmatrix}-I_2, & I_2\end{bmatrix}$, and $\boldsymbol{\theta} = (0.5, 1, 2, 2)$. For $p = 2$, let $\tilde{M} = \begin{bmatrix}-I_3, & I_3\end{bmatrix}$, and $\boldsymbol{\theta} = (0.5, 0.5, 1, 2, 1, 2)$. For $p = 3$, let $\tilde{M} = \begin{bmatrix}-I_4, & I_4\end{bmatrix}$, and $\boldsymbol{\theta} = (0.5, 0.5, 0, 1, 2, 1, 0, 2)$. Let $\tilde{M}\boldsymbol{\theta} = r$ be the restrictions. To calculate Δ, let $r_0 = 0.75kr$, $k = 1, 2, 3, 4, 5, 6$ and to highlight the performance of the proposed test, in Figure 2.15-Figure 2.17 the variation of the empirical power versus the noncentrality parameter Δ under different significant level with $T = 20$ are reported. Figure 2.15- Figure 2.17

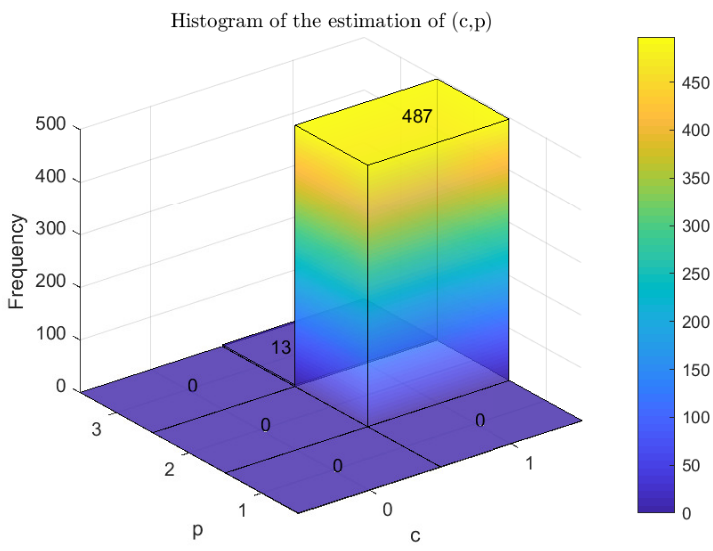

Figure 2.13: The histogram of the estimator of $(\hat{c}, \hat{p})$ $(T = 35)$

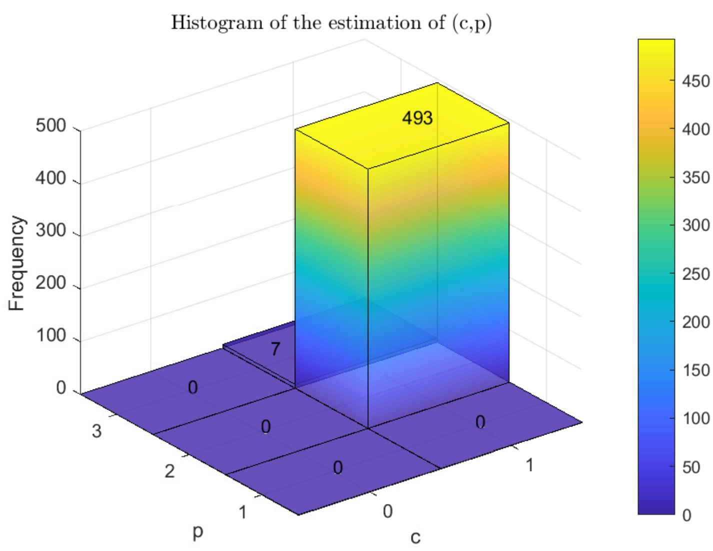

Figure 2.14: The histogram of the estimator of $(\hat{c}, \hat{p})$ $(T = 50)$

indicate that the power of the test is the highest when p takes its exact value, i.e. if $p = 2$. Figure 2.15- Figure 2.17 also show that the empirical power tends to 1 as Δ

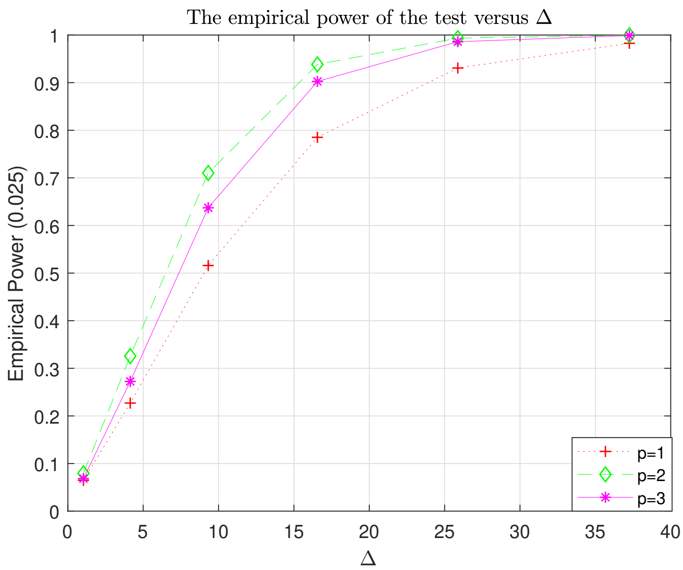

Figure 2.15: The empirical power of the test with different Δ and p ($T = 20, \alpha = 0.025$)

increases to infinity. All figures show that, as T increases to infinity, the empirical power also increases to 1 and they confirm that the proposed test is consistent. Further, from Figure 2.15- Figure 2.17, the empirical power takes the maximum value at $p = 2$.

Furthermore, the performance of the UMLE, the RMLE and the SEs are evaluated. By 500 replications, calculate the mean squared error of each estimator according to different non-centrality parameter with a nonnegative weighting matrix $\Omega = I_6$. To specify the restriction (2.3.14), let $\tilde{M} = \begin{bmatrix} I_3, & -I_3 \end{bmatrix}$, and $r = \tilde{M}\boldsymbol{\theta}$, where the pre-assigned parameter $\boldsymbol{\theta} = (0.5, 0.5, 1, 2, 1, 2)^\top$. Evaluate the relative mean squared efficiency (RMSE) of each estimator which is given by

$$\text{RMSE}(\hat{\theta}_0) = \text{ADR}(\hat{\theta}_T, \theta, \Omega)/\text{ADR}(\hat{\theta}_0, \theta, \Omega) \tag{2.6.2}$$

where $\hat{\theta}_0$ represents an estimator such as $\tilde{\theta}_T$, $\hat{\theta}_T^{sh}$, $\hat{\theta}_T^{sh+}$ and $\hat{\theta}_T$. To generate the data under the alternative hypothesis (2.3.23), let $r_0 = 0.75kr$, $k = 1, 2, 3, 4, 5, 6$. Compute Δ as

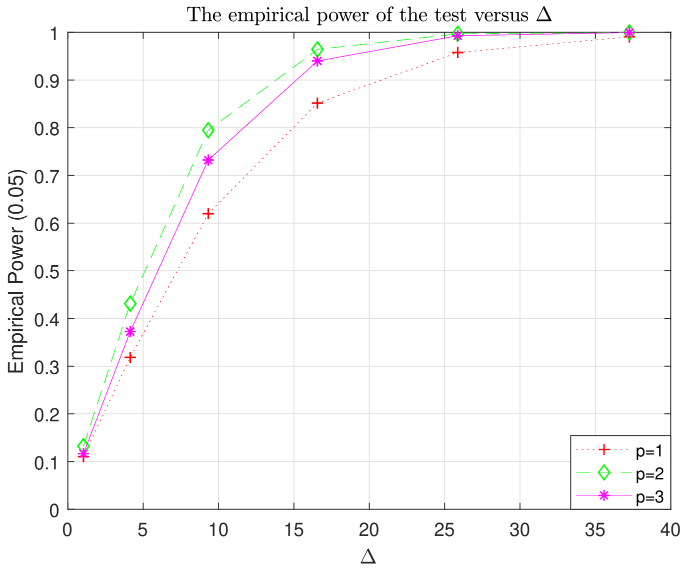

Figure 2.16: The empirical power of the test with different Δ and p ($T = 20, \alpha = 0.05$)

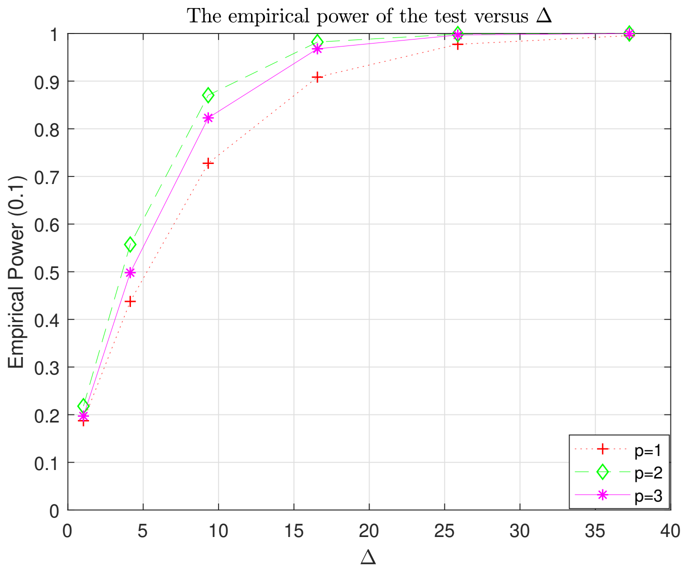

Figure 2.17: The empirical power of the test with different Δ and p ($T = 20, \alpha = 0.1$)

$\Delta = \frac{1}{\sigma^2} r_0^\top (\tilde{M}\mathbf{\Sigma}^{-1}\tilde{M}^\top)^{-1} r_0$. From Figure 3.2 to Figure 3.3, near $\Delta = 0$, RMLE has the best performance, which means that near the null hypothesis, RMLE is more efficient than the UMLE, SEs and PSE. These figures also indicate that the efficiency of RMLE decreases as one moves far away from the restriction. Its performance tends to be the worst as Δ tends to infinity. Furthermore, such figures also show that PSE is always more efficient than SEs, which confirms Proposition 2.5.2. Meanwhile, RMSE of both SEs and PSE are decreasing as Δ is far away from the origin, but they are always greater than 1. In Appendix A.1, the relative efficiency of the proposed estimators for the cases where $T = 50$ and $T = 80$ is reported. The portray of these figures is similar to that of Figure 2.18-Figure 2.21.

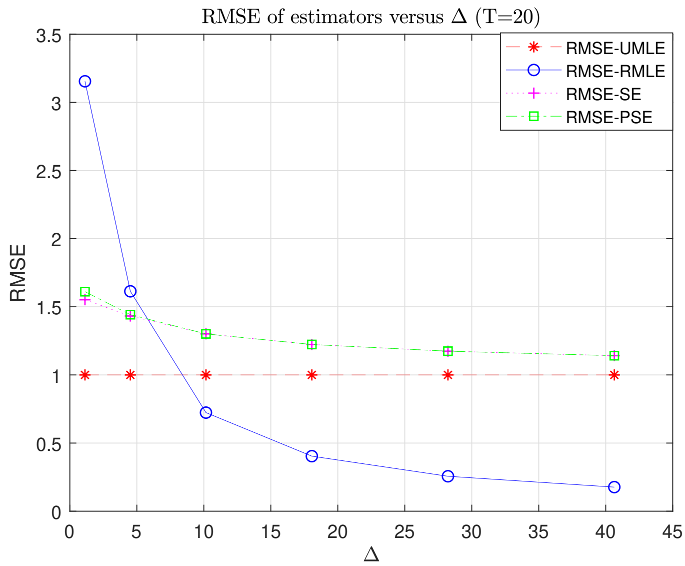

Figure 2.18: The RMSE of the estimator versus Δ ($T = 20$)

2.6.3 Real data set: financial market data

In this subsection, the proposed method is applied to the gold spot daily price dataset for the period from Dec 29, 1978 to Jan 12, 2022. The data set is available at https://goldprice.org and https://macrotrends.dpdcart.com. The proposed method is

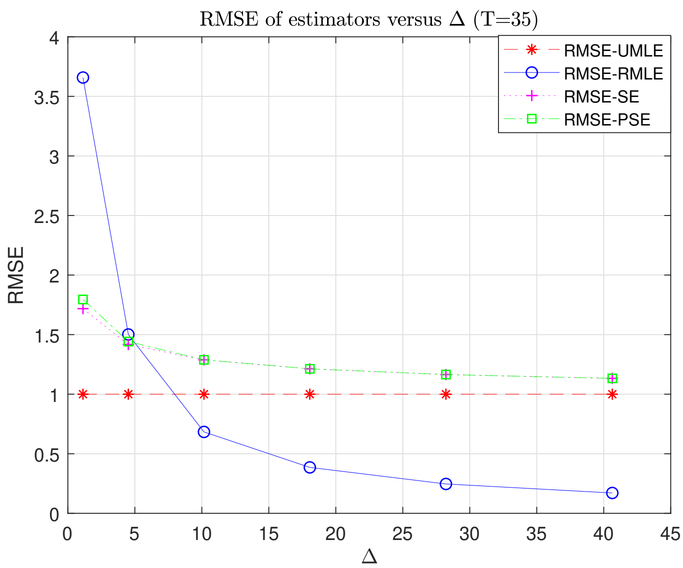

Figure 2.19: The RMSE of the estimator versus Δ ($T = 35$)

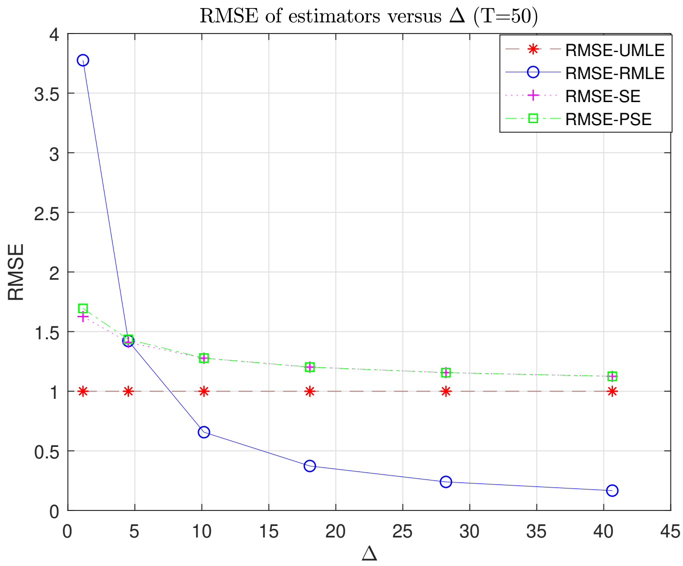

Figure 2.20: The RMSE of the estimator versus Δ ($T = 50$)

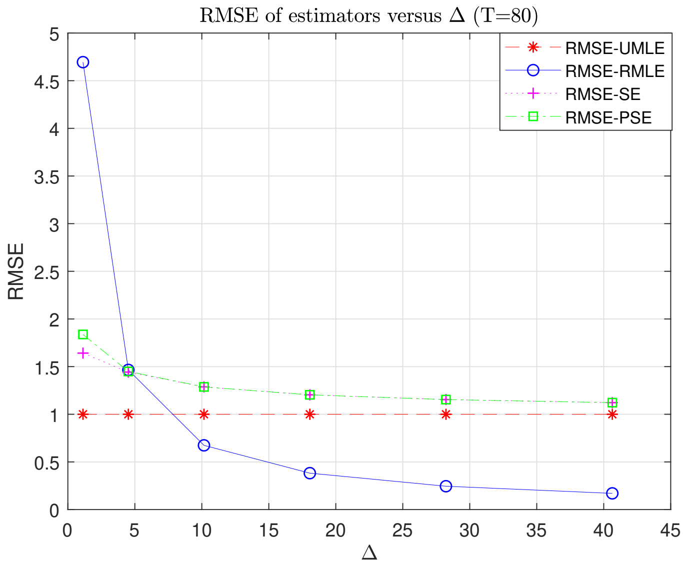

Figure 2.21: The RMSE of the estimator versus Δ ($T = 80$)

used to detect the existence of change-point t^* and to estimate p. The horizon period time is taken as $T = 43$ and the total number of records is $N = 11175$, which gives $\Delta = T/N \approx 0.0038$. As shown in empirical studies of Schwartz [1997] and Chen [2010], the mean-reversion features holds for prices of several commodities including oil, gold, gas, corn etc... Thus, a GEOU-type model is chosen to fit such a dataset. The prices are fitted with several different models, for example, with one change-point and different value of p. To apply the proposed method, σ is estimated by the data's realised volatility $\left(\hat{\sigma} = \sqrt{\sum_{i \in \mathbb{N}[0,T]} (\ln X(t_{i+1}) - \ln X(t_i))^2 / T} = 0.1919\right)$. If p is odd, use the model

$$dX(t) = \sum_{j=1}^{2} \left[\sum_{i=1(2)}^{p} \left(\mu_{ij} \sqrt{2} \cos\left(2\pi \frac{i-1}{2} t\right) + \mu_{(i+1)j} \sqrt{2} \sin\left(2\pi \frac{i-1}{2} t\right) - \alpha_j \ln X(t) \right) \right] \times X(t) \mathbb{I}_{\{t_{j-1}^* \leqslant t \leqslant t_j^*\}} dt + \sigma X(t) dW_t. \tag{2.6.3}$$

If p is even, the model is stated as

$$dX(t) = \sum_{j=1}^{2} \left(\left(\sum_{i=1(2)}^{p-1} \left[\mu_{ij} \sqrt{2} \cos\left(2\pi \frac{i-1}{2} t\right) + \mu_{(i+1)j} \sqrt{2} \sin\left(2\pi \frac{i-1}{2} t\right) \right] \right.\right.$$

$$+\mu_{pj}\sqrt{2}\cos\left(2\pi\frac{p}{2}t\right)\Big) - \alpha_j \ln X(t)\Big)X(t)\mathbb{I}_{\{t^*_{j-1}\leqslant t\leqslant t^*_j\}}dt + \sigma X(t)dW_t. \quad (2.6.4)$$

where $\sum_{i=1(2)}^{p}$, $\sum_{i=1(2)}^{p-1}$ stand for taking summation with step 2 and $c = 0, 1$. In practice, suppose that $p_{\max} = 5$. The proposed method is applied to estimate c, p as well as the UMLE, the RMLE and the SEs. To set up the restriction, the case where the mean reversion level remains the same before and after the change point is considered. Note that, detecting the existence of the change-point leads to the hypothesis testing problem: $H_0 : M\theta = 0$ versus $H_1 : M\theta \neq 0$, where M is given in Table 2.11. The test statistics and critical value at different significant level are reported in Table 2.10. This table shows that that, at significance level 0.1, the test statistics fall into the rejection region for all p. Thus, the null hypothesis is rejected at significance level $\alpha = 0.1$ for all $p = 1, 2, 3, 4, 5$. *Bootstrap method* on residuals is used to analyse the relative efficiency

Table 2.10: Test statistics value and critical value

(c,p)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)
Critical value ($\alpha = 0.1$)	4.6052	6.2514	7.7794	9.2364	10.6446
Critical value ($\alpha = 0.05$)	5.9915	7.8147	9.4877	11.0705	12.5916
Critical value ($\alpha = 0.025$)	7.3778	9.3484	11.1433	12.8325	14.4494
Statistics value	6.8596	8.1440	9.3628	12.3800	12.8755

Table 2.11: Power of test

(c,p)	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)
M	$[I_2, -I_2]$	$[I_3, -I_3]$	$[I_4, -I_4]$	$[I_5, -I_5]$	$[I_6, -I_6]$
$M\theta$	$\begin{bmatrix}0.3943\\0.0061\end{bmatrix}$	$\begin{bmatrix}0.5291\\0.0214\\0.0250\end{bmatrix}$	$\begin{bmatrix}0.5065\\0.0194\\0.0538\\0.0218\end{bmatrix}$	$\begin{bmatrix}0.5171\\0.0212\\0.0501\\0.0876\\0.0216\end{bmatrix}$	$\begin{bmatrix}0.7224\\0.0190\\0.0512\\0.0831\\0.0266\\0.0520\end{bmatrix}$
power ($\alpha = 0.1$)	0.6040	0.5690	0.5280	0.6370	0.6060

and the power function based on 1000 replicates. Namely, the Bootstrapped RMSE and the empirical power of the test are calculated. Table 2.11 shows the results obtained for the empirical power at the significant level $\alpha = 0.1$. From this table, one can see that when $c = 1, p = 4$, the power of the test reaches its highest value. This indicates that the GEOU process in (2.6.4) with $c = 1, p = 4$ is the most suitable one to fit the *daily gold spot historical prices*. Thus, the data series can be fitted by the following model

$$\begin{aligned} dX(t) = &\Big(\Big(\mu_{11} + \mu_{12}\sqrt{2}\cos(2\pi t) + \mu_{13}\sqrt{2}\sin(2\pi t) + \mu_{14}\sqrt{2}\cos(4\pi t)\Big) \\ &-\alpha_1 \ln X(t)\Big) X(t)\mathbb{I}_{\{0\leqslant t\leqslant t^*\}}dt + \Big(\Big(\mu_{21} + \mu_{22}\sqrt{2}\cos(2\pi t) + \mu_{23}\sqrt{2}\sin(2\pi t) \\ &+\mu_{24}\sqrt{2}\cos(4\pi t)\Big) - \alpha_2 \ln X(t)\Big) X(t)\mathbb{I}_{\{t^*\leqslant t\leqslant T\}}dt + \sigma X(t)dB_t. \end{aligned} \tag{2.6.5}$$

From Table 2.11, it can also be noticed that increasing the value of p does not always result in increasing the empirical power. For example, the power at $c = 1, p = 4$ is greater than the power at the point $c = 1, p = 5$. For the linear restriction, the matrix M defined as in Table 2.11 is used. The result is reported in Table 2.12. For $c = 1$ and $p =$

Table 2.12: Parameter estimation for real data

	μ_{11}	μ_{12}	μ_{13}	μ_{14}	α_1	μ_{21}	μ_{22}	μ_{23}	μ_{24}	α_2
UMLE	2.3101	0.0211	0.0097	0.0669	0.3258	1.6987	0.0020	-0.0417	-0.0182	0.2887
RMLE	0.3506	0.0048	-0.0277	0.0067	0.0600	0.3506	0.0048	-0.0277	0.0067	0.0600
SEs	1.7407	0.0165	-0.0010	0.0500	0.2485	1.3258	0.0028	-0.0377	-0.0122	0.2254
PSE	1.7525	0.0166	-0.0008	0.0502	0.2501	1.3328	0.0027	-0.0378	-0.0113	0.2266

4, the (bootstrapped) RMSE of 0.7201 is obtained, 1.2510 and 1.2924 for the RMLE, the SEs and the PSE, respectively are also obtained. Thus, the bootstrapped RMSE are in agreement with the theoretical results for which the SEs and PSE dominate the UMLE and RMSE has the worst performance under the condition that null hypothesis $M\theta = 0$ is strongly rejected. Further, to predict the daily gold spot Historical Price from December 29, 1978 to January 12, 2022, the model (2.6.5) with $p = 4$ is used and the UMLE given in Table 2.12. Figure 2.22 gives a graph representative of the real price data and the predicted price based on the UMLE. From Figure 2.22, one can see that the predicted data reflects the basic trend of the real price data. The predicted prices based on the RMLE and the SEs give similar graphs. Because of that and to save the space of this chapter, these graphs are not reported here.

Comparing with other cited works which analyze the *daily gold spot historical*

Figure 2.22: Daily gold spot price from Dec 29, 1978 to Jan 12, 2022 and fitted price

prices, the novelty of this chapter consists in the fact that the dimensions of the drift parameter of the GEOU is estimated, and a test is proposed to detect the number of change-point. Furthermore, by bootstrap method, the relative efficiency of the UMLE, the RMLE and the SEs are evaluated. Finally, the portray given by Figure 2.22 is in agreement with the established theoretical result, given in Proposition 2.4.6, which indicates that the proposed method improves the goodness-of-fit.

2.7 Conclusion

In this chapter, a GEOU process that works well for positive financial datasets that have a periodic mean-reverting level was proposed. In passing, it should be stressed that many financial datasets take only positive values. In comparison with other existing works about the exponential O-U processes, the proposed GEOU is incorporate the seasonality effect. On top of that flexibility, the dataset under consideration may be subject to a drastic change. A statistical method was proposed, which can be used to validate or not the seasonality effect or the drastic change as well as other possible relations binding the components of the drift parameter. To this end, the UMLE, RMLE, and their joint asymptotic normality as well as the strong consistency of the UMLE are . Based on these asymptotic results, a test was constructed for testing a restriction about the drift parameter. Further, in contrast with the works in recent cited literature,the case where the component of the drift parameter vector is unknown was considered. Based on a *Schwartz Information Criterion*, a statistical method which is useful in estimating the dimension of the drift parameter as well as the change-point and the drift parameter was presented. A class of shrinkage type estimators was derived, which encloses as special cases the UMLE and the RMLE. Thanks to the ADR, the relative performance of the RMLE and that of the SEs as compared to the UMLE were established. In particular, it is established that, the SEs dominate the UMLE and that, near the null hypothesis, the RMLE is the most efficient. However, the RMLE is dominated by the UMLE as one moves far away from the restriction. These theoretical findings are confirmed by the simulation studies and in order to illustrate the application of the proposed method, the *daily gold spot historical prices* was analyzed. Finally, the proposed method improves the goodness-of-fit and this theoretical result is confirmed via the fitted values of the *daily gold spot historical prices* which are very close to the observed datasets.

Chapter 3

Inference in GCIR Process

3.1 Introduction

Over several years, there has been a growing interest in using some mean-reverting processes in order to model some financial data see Vasicek [1977], economical data Langetieg [1980], physical phenomena Lansky and Sacerdote [2001], biological phenomena Rohlfs et al. [2010] among others. To give some references, see Vasicek [1977], Langetieg [1980], Lansky and Sacerdote [2001], Rohlfs et al. [2010] and references therein. The mean-reverting process used in the above quoted papers is known as Ornstein-Uhlenbeck process or Vasicek process. As an extension of the Vasicek model, John C. Cox and Ross [1985a,b] introduced, in 1985, a stochastic process known as Cox–Ingersoll–Ross (CIR) model. The proposed CIR model is given by the stochastic differential equation (SDE)

$$dx(t) = \alpha(\beta - x(t))dt + \sigma\sqrt{x(t)}dB_t, \quad x_0 > 0, \tag{3.1.1}$$

where $\{B_t, t \geqslant 0\}$ is a Brownian motion (modelling the random market risk factor) and β, α, σ, are parameters. The parameter α represents the speed of adjustment to the mean β, and σ is the volatility. The standard deviation factor, $\sigma\sqrt{x(t)}$, prevents the possibility of negative interest rates for all positive values of β and α. An interest rate of zero is also precluded if the condition $2\alpha\beta \geqslant \sigma^2$ holds. The quantity $4\alpha\beta/\sigma^2$ plays a critical role in the behavior of the process. It is known as the *dimension* of the process $\{x(t), t \geqslant 0\}$. The CIR model (3.1.1) is a type of "one-factor model" as it describes interest rate movements driven by only one source of market risk. The model (3.1.1) can be used in the valuation of interest rate derivatives. For About three decades, the parameter estimation in CIR and the properties of the CIR processes have received considerable attention. For instance, Chen and Scott [1993] extended the single-factor equilibrium model to a multi-factor setting and estimated the parameters by the maximum likelihood method. Another interesting reference is Maghsoodi [1996] who generalized the CIR model to the case of time-varying parameters and proved the trajectory can be viewed as a log-normal process through a stochastic time change. About two decades later, Maboulou and Mashele [2015] estimated the parameters for multi-factor affine CIR-type hazard rate model. Another citation is Feng and Xie [2012] who considered the Bayesian estimation of interest rate model (3.1.1) based on Euler-Maruyama approximation. Further, see Alaya and Kebaier [2012, 2013], combined the two who dealt with the problem of global parameter estimation in the CIR model, and derived the distribution of the estimators. To give another interesting reference, see the citation Peng and Schellhorn [2018] who proved that the distribution of generalized CIR process with time-varying parameters can be represented as a convergent series of weighted independent central and non-central chi-square random variables. Just recently, Tong and Zhang [2017], Zhang et al. [2019, 2020], combined the two Tong et al. [2021], considered a type of CIR interest rate model with random switching and stated the sufficient conditions for the ergodicity of the solution.

With respect to other stochastic processes, the classical CIR model possesses the main characteristic of the Ornstein–Uhlenbeck process, i.e. the tendency to return towards the long-term equilibrium. This property is known as *mean-reversion*. As pointed out in Geman [2009], such property is found in several applications including commodity and energy price processes. However, despite such a general trend, the assumption of a constant mean-reversion level seems inadequate due to seasonality patterns or a long-term trend of the process.

This chapter considers a more general process which incorporates a deterministic and periodic drift term in the SDE. In particular, the inference problem about the drift parameters, in the context where some prior information (from outside the sample) may be available is considered. For example, the source of prior information may be the expertise in a certain field, which establishes an association among the parameters to be estimated. Another source of prior information may be the previous statistical investigations which may have established that there exists a linear restriction binding the drift parameter and some known column vectors. For more details on the source of uncertainty about the prior information in linear models is referred to Nkurunziza [2015] and references. In such a case, unrestricted maximum likelihood estimator (UMLE) may not be optimal. Thus, it is interesting to derive a statistical method which combines the prior information and the sample information. Further, to overcome some uncertainty about the restriction, it is interesting to derive a test for testing the hypothesized restriction.

The rest of this chapter is structured as follows. Section 3.2 presents the statistical model and some preliminary results as well as some useful properties of the trajectory of the proposed GCIR process. In Section 3.2, an approximate auxiliary process which is strictly stationary and ergodic is introduced. This helps derive the asymptotic properties of the proposed estimators. Section 3.3 derives the UMLE and the restricted maximum likelihood estimator (RMLE) of the drift parameter. In this section, the joint asymptotic normality of the UMLE and RMLE, under the set of local alternative restrictions is also established. In Section Section 3.4, a test for testing the hypothesized restriction is derived and SEs are proposed. Section 3.5 establishes the asymptotic distributional risk (ADR) of the proposed estimators and by using the ADR. The asymptotic dominance of these estimators is also studied in this section. Section 3.6 presents the empirical study. Finally, to illustrate the application of the proposed method, a data analysis of the historical corn price as well as that of U.S. 10-Year Treasury Bond Yield historical data is performed. For more details, we refer to Lyu and Nkurunziza [2023a].

3.2 Statistical model and preliminary results

3.2.1 Statistical model

Strongly inspired from the work in Dehling et al. [2010] and Nkurunziza and Zhang [2018], the statistical model under consideration is a generalization of CIR process (3.1.1). Specifically, consider observing a stochastic process $\{X(t), t \geqslant 0\}$ which is a solution of the SDE

$$dX(t) = S(\theta, t, X(t))dt + \sigma \sqrt{X(t)}dB_t, \quad X(0) = X_0, \tag{3.2.1}$$

where $S(\theta, t, X(t)) = L(t) - \alpha X(t)$, with $L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t)$. In the sequel, the process in (3.2.1) will be referred to as the GCIR process (or GCIR model). For the special case where $L(t)$ is a constant, note a slight difference between the original CIR process (3.1.1) and model (3.2.1) in the position of α within the drift term. Nevertheless, the GCIR model (3.2.1) is transformed to a process with drift term $\alpha(\widetilde{L(t)} - X(t))dt$ where $\widetilde{L(t)} = L(t)/\alpha$. The advantage of (3.2.1) compared with the process provided with the drift $\alpha(\widetilde{L(t)} - X(t))dt$ is the simplification of the study of the estimators. Thus, the parametrization in (3.2.1) is considered here for the sake of simplicity. In this chapter, θ is the parameter of interest. Note that, in the continuous time observations, the diffusion parameter σ^2 can be consistently estimated by the discretized version of quadratic variation of the process $\left\{2\sqrt{X(t)}, t \geqslant 0\right\}$. Because of that, this chapter assumes that σ^2 is known. The inference is performed under the situation where there may exist prior information on the target parameter. In particular, estimation problem about θ is studied under the context where the target parameter may satisfy the linear restriction (2.2.2). As stated in Nkurunziza and Zhang [2018], the above restriction indicates that there exists some linear relation binding some components of the drift parameter vector. A statistical method is developed, which preserve a good performance whenever the restriction is valid or not. Further, note that the restriction (2.2.2) leads to the hypothesis testing problem (2.2.3) Thus, a test is derived for testing the restriction in (2.2.3). The optimality of the proposed method is based on the established asymptotic properties of the UMLE and the RMLE. The derivation of the established main results relies on Assumption 2.1, Assumption 2.2 and the following assumptions.

Assumption 3.1. The distribution of the initial value, X_0, of the SDE in (3.2.1) does not depend on the drift parameter θ. Further, $X_0 \geqslant x_0 \geqslant 0$ a.s. where x_0 is the initial value of the SDE (3.1.1). X_0 is independent of $\{B_t : t \geqslant 0\}$ and $\mathbb{E}(|X_0|^m) < \infty$, for some $m \geqslant 2$.

First, to prevent the process $\{X(t), t \geqslant 0\}$ from turning negative, some restrictions on the parameters and the mean-reverting term $L(t)$ of this GCIR model are needed.

Assumption 3.2. The function $L(t) \geqslant \max\{\alpha\beta, \sigma^2\}$ with $4\alpha\beta/\sigma^2 \geqslant 2$.

It should be noticed that, since the function $\varphi(t)$ is Riemann-integrable on $[0, T]$ and υ-periodic, Assumption 2.2 implies that $\varphi(t)$ is bounded on $\mathbb{R}_+ = [0, +\infty)$. As in Dehling et al. [2010], p is supposed to be known as well as the function $\varphi(t)$. Without loss of generality, suppose that the period υ is known and equals to 1. Proposition B.2 and [Karatzas and Shreve, 1998, Proposition 2.18] are used to compare the solution of the classical CIR process and the one of the GCIR process (3.2.1). The role of Assumption 3.2 is to guarantee that the conditions of Proposition B.2 hold.

Remark 3.2.1. Let $d(t) = 4L(t)/\sigma^2$. $d(t)$ is known as the dimension of the GCIR model (3.2.1), which plays an important role in the behavior of the process $\{X(t), t \geqslant 0\}$.

Under these assumptions, in the next subsection, the existence and uniqueness of the strong solution of the GCIR process (3.2.1) is given. Note that the major difficulty of the problem studied consists in the fact that the SDE in (3.2.1) does not have an explicit solution expression. However, the implicit solution can be obtained, which allows us to determine the L^m boundedness of this solution. An auxiliary process, which is stationary and ergodic, is constructed. Furthermore, it is proven that the distance between the strong solution and the auxiliary process converges, in L^1 and almost surely, to 0.

3.2.2 Existence of strong and unique solution

This subsection aims at deriving the existence and uniqueness of a nonnegative strong solution of the process (3.2.1). These properties can help us to study the upper and lower bounds of this solution. This is an important step in obtaining an explicit expression of the likelihood function. The following proposition gives the existence and uniqueness of the nonnegative strong solution.

Proposition 3.2.1. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then, the GCIR model in* (3.2.1) *admits a strong and unique non-negative solution on* $[0, T]$, *for* $0 \leqslant T < \infty$.

The proof of this proposition is given in Appendix B.1. From Theorem 7.6 of Liptser and Shiryaev [2001], to get the likelihood function of a diffusion process, it is sufficient to guarantee that $\mathrm{P}\left(\int_0^T S^2(\theta, t, X(t))\big/(\sigma^2 X(t))dt < \infty\right) = 1$, for all $0 \leqslant T < \infty$, and for all $\theta \in \Theta$. The following corollary guarantees that the sufficient condition for the likelihood function holds.

Corollary 3.2.1. *If Assumption 2.1-2.2, , and Assumption 3.1-3.2 hold, then*
$\mathrm{P}\left(\int_0^T \frac{S^2(\theta, t, X(t))}{\sigma^2 X(t)} dt < \infty\right) = 1,\ \forall 0 \leqslant T < \infty,\ \forall \theta \in \Theta.$

The proof is given in the Appendix B.1. The following result gives an implicit form of the solution of the SDE (3.2.1). The established result gives also the expectation and variance of the solution of the model (3.2.1). The methodological approach taken in Proposition 3.2.2 is a mixed methodology based on Itô's formula.

Proposition 3.2.2. *Under Assumption 2.1-2.2, , and Assumption 3.1-3.2,*

(1) *the solution of the SDE (3.2.1) can be rewritten as*

$$X(t) = e^{-\alpha t} X_0 + h(t) + Z(t), \quad t \geqslant 0, \tag{3.2.2}$$

where $h(t) = e^{-\alpha t} \int_0^t e^{\alpha s} L(s) ds$, $Z(t) = \sigma e^{-\alpha t} \int_0^t e^{\alpha s} \sqrt{X_s} dB_s$.

(2) $\mathbb{E}[X(t)] = e^{-\alpha t}\mathbb{E}[X_0] + \sum_{i=1}^{p} \mu_i \int_0^t e^{-\alpha(t-s)} \varphi_i(s) ds,$

(3)$\mathrm{Var}(X(t)) = e^{-2\alpha t}\left(\mathrm{Var}(X_0) + \sigma^2 \mathbb{E}[X_0]\frac{1}{\alpha}(e^{\alpha t} - 1) + \sigma^2 \int_0^t e^{\alpha s} \int_0^s e^{\alpha u} L(u) du ds\right).$

Further, $X(t)$ *is* L^2-*bounded, i.e.* $\sup_{t \geqslant 0} \mathbb{E}[X^2(t)] < \infty$.

The proof of this proposition is given in Appendix B.1. From Part **(3)** of Proposition 3.2.2, the following proposition proves that the solution $\{X(t), t \geqslant 0\}$ is L^m bounded. To this end, let $\sum_{i=1}^{p} |\mu_i| \leqslant K_\mu$ and $\|\varphi(t)\| \leqslant K_\varphi$, for some positive constants K_μ, K_φ, then,

$$0 < L(t) \leqslant K_\varphi K_\mu. \tag{3.2.3}$$

Further, let

$$K = \mathbb{E}\left[X_0^2\right] + \frac{K_\varphi^2 K_\mu^2}{\alpha^2} + \sigma^2 \mathbb{E}[X_0]\frac{1}{4\alpha} + \sigma^2 K_\varphi K_\mu \left(\frac{1}{2\alpha^2} - \frac{1}{4\alpha^2}\right) + K_\varphi K_\mu \frac{1}{2\alpha}.$$

By Assumption 3.1, $\mathbb{E}\left[X_0^2\right] < \infty$, then,

$$\sup_{t \geqslant 0} \mathbb{E}[X(t)^2] \leqslant K. \tag{3.2.4}$$

Moreover, let C_m be some positive constant that only depends on m, and let

$$K_m \quad = \mathbb{E}[X_0^m] + C_m \mathbb{E}[X_0^{m-1}]\tfrac{1}{\alpha m} + C_m C_{m-1} \mathbb{E}[X_0^{m-2}]\tfrac{1}{\alpha(m-1)}\tfrac{1}{\alpha m} + \cdots$$

$$+\max\{1,\mathbb{E}[X(t)]\}C_mC_{m-1}\cdots C_{m-(k-1)}\mathbb{E}[X_0^{m-(k-1)}]\tfrac{1}{\alpha(m-(k-1))}\cdots\tfrac{1}{\alpha m}. \tag{3.2.5}$$

Proposition 3.2.3. *If Assumption 2.1-2.2, , and Assumption 3.1-3.2 hold, then,*

$$\sup_{t\geqslant 0}\mathbb{E}[X(t)^m]\leqslant K_m. \tag{3.2.6}$$

The proof of this proposition is given in the Appendix B.1.

Remark 3.2.2. By Jensen's inequality, it is obvious that

$$\mathbb{E}\left[1/X(t)\right]\geqslant 1/\mathbb{E}\left[X(t)\right]=1/\left(e^{-\alpha t}\mathbb{E}[X_0]+h(t)\right).$$

One of the main challenge to overcome consists in the fact that the process $\{X(t),t\geqslant 0\}$ is not stationary except in the special case where the dimension of the GCIR process is a positive integer. Moreover, the solution $\{X(t),t\geqslant 0\}$ has no explicit expression. To overcome this difficulty, below, an auxiliary process which is strictly stationary and ergodic is constructed. Furthermore, it is proven that the distance between the auxiliary process and the solution of the SDE (3.2.1) converges to 0 both in L^1 and almost surely. The convergence allows us to derive the asymptotic distributions of the UMLE and RMLE by using the ergodicity of the auxiliary process. The limiting distributions of both UMLE and RMLE play important roles in testing the restrictions. Suppose that there exists one process $X(t)^*=(X^{(1*)}(t),X^{(2*)}(t))$, satisfying the following SDEs:

$$dX^{(1*)}(t)=\left(\frac{\sigma^2}{2}-\alpha X^{(1*)}(t)\right)dt+\sigma\sqrt{X^{(1*)}(t)}dB_t^{(1*)},\ X^{(1*)}(0)=\frac{X_0}{2}, \tag{3.2.7}$$

$$dX^{(2*)}(t)=\left(L(t)-\frac{\sigma^2}{2}-\alpha X^{(2*)}(t)\right)dt+\sigma\sqrt{X(t)^{(2*)}}dB_t^{(2*)},\ X^{(2*)}(0)=\frac{X_0}{2}, \tag{3.2.8}$$

where $B_t^{(*)}=(B_t^{(1*)},B_t^{(2*)})$ are two dimensional standard Brownian motions defined on the probability space $(\Omega,\mathscr{F},\mathrm{P})$. Proposition 3.2.1 indicates that SDE (3.2.7) and (3.2.8) admit strong solutions. The following proposition states the relation between the solutions $X^{(1*)}(t)$, $X^{(2*)}(t)$ and $X(t)$, where $X(t)$ is the solution to SDE (3.2.1).

Proposition 3.2.4. *Let $X^{(1*)}(t)$ be the strong solution to SDE (3.2.7) and let $X^{(2*)}(t)$ be the strong solution to SDE (3.2.8). Then, for $t \geqslant 0$, $X^{(1*)}(t) + X^{(2*)}(t)$is the strong solution of SDE (3.2.1), i.e. $X^{(1*)}(t) + X^{(2*)}(t) = X(t)$ almost surely.*

The proof of this proposition is given in the Appendix B.2. The dimension of SDE (3.2.8) may be less than 2, the following result shows that its solution is positive with probability 1.

Proposition 3.2.5. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then, for all $w \in \Omega$, $\int_0^T \mathbb{I}_{A_t}(w)dt = 0$ a.s. for $0 \leqslant T < \infty$, where $A_t(w) = \{\omega : X^{(2*)}(t, w) = 0, t \geqslant 0\}$.*

The proof is similar to the proof of Lemma 2.2 in Tong and Zhang [2017] and given in Appendix B.2. The following proposition is useful in constructing a stationary and ergodic auxiliary process.

Proposition 3.2.6. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then, the solution of SDE (3.2.7) is given as $X^{(1*)}(t) = \sum_{j=1}^{2} Y_j^2(t)$, where for $j = 1, 2$, $Y_j(t) = e^{-\frac{\alpha}{2}t}Y_j(0) + \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}dB_j(s)$. Further, let*

$$Y_t = e^{-\frac{\alpha}{2}t}\sqrt{X_0^{(2*)}} + e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}\left(4L(s) - 3\sigma^2\right)/\left(8\sqrt{X(t)^{(2*)}}\right)ds + \frac{\sigma}{2}e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}dB_s^{(2*)},$$

the solution of SDE (3.2.8) can be rewritten as $X^{(2*)}(t) = Y_t^2$.

The proof of this proposition is given in Appendix B.2.

3.3 The unrestricted and restricted estimators

In this section, the UMLE and RMLE are derived. Further, the joint asymptotic normality of the UMLE and RMLE is established.

3.3.1 The UMLE $\hat{\theta}_T$ and the RMLE $\tilde{\theta}_T$

Let $\mathrm{P}^{(\theta)}_{X^T}$ denote the measure induced by the observable realizations $X^T = \{X(t), t \geqslant 0\}$ on the measurable space $(C_{[0,T]}, \mathcal{B}_{[0,T]})$, where $C_{[0,T]}$ is the space of continuous, real-valued functions on $[0,T]$ and $\mathcal{B}_{[0,T]}$ is the associated Borel σ-algebra. Further, let P_B be the measure generated by the Brownian motion on $(C_{[0,T]}, \mathcal{B}_{[0,T]})$. Then, the Radon-Nikodym derivative of observations X^T is given by $\mathcal{L}(\theta, X^T) = d\mathrm{P}^{(\theta)}_{X^T}/d\mathrm{P}_B(X^T)$. Thus, the UMLE can be derived by minimizing the functional $\theta \longmapsto \mathcal{L}(\theta, X^T)$, i.e. $\hat{\theta}_T = \arg\max_{\theta} \mathcal{L}(\theta, X^T)$. To simplify some notations, let

$$Q_{[0,T]} = \begin{bmatrix} \displaystyle\int_0^T \frac{\varphi^\top(t)\varphi(t)}{X(t)}dt & -\displaystyle\int_0^T \varphi^\top(t)dt \\ -\displaystyle\int_0^T \varphi(t)dt & \displaystyle\int_0^T X(t)dt \end{bmatrix}_{(p+1)\times(p+1)}, \tag{3.3.1}$$

$$\begin{aligned} R_{[0,T]} &= \left(\int_0^T \frac{\varphi(t)}{X(t)}dX(t), -\int_0^T \frac{X(t)}{X(t)}dX(t)\right)^\top, \\ W_{[0,T]} &= \left(\int_0^T \frac{\varphi(t)}{\sqrt{X(t)}}dB_t, -\int_0^T \frac{X(t)}{\sqrt{X(t)}}dB_t\right)^\top. \end{aligned} \tag{3.3.2}$$

For the purpose of minimizing the functional $\mathcal{L}(\theta, X^T)$, and deriving the UMLE and RMLE, the matrix $Q_{[0,T]}$ needs to be invertible. The following proposition proves that the matrix $Q_{[0,T]}$ is positive definite provided that $T \geqslant 1$.

Proposition 3.3.1. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then, $Q_{[0,T]}$ is a positive definite matrix whenever $T \geqslant 1$.*

The proof of this proposition is given in Appendix B.3. Because the optimality of the proposed method is asymptotic, in the sequel, without loss of generality, the condition $T \geqslant 1$ is always supposed to hold. Let $\tilde{\theta}_T$ be the RMLE and let $G_{[0,T]} = Q^{-1}_{[0,T]} M^\top (M Q^{-1}_{[0,T]} M^\top)^{-1}$.

Proposition 3.3.2. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then, $\hat{\theta}_T = Q^{-1}_{[0,T]} R_{[0,T]}$, and $\tilde{\theta}_T = \hat{\theta}_T - G_{[0,T]}(M\hat{\theta}_T - r)$.*

The proof of this proposition is given in Appendix B.3.

3.3.2 Auxiliary process

To establish the auxiliary process, for $j = 1, 2$, let $\{B_j(s), s \geqslant 0\}$ and $\{\bar{B}_j(s), s \geqslant 0\}$ be two independent Brownian motions and $\tilde{B}_j(s)$ be a bilateral Brownian motion with $\tilde{B}_j(s) = B_j(s)\mathbb{I}_{\{s>0\}} + \bar{B}_j(-s)\mathbb{I}_{\{s\leqslant 0\}}$. Let

$$\begin{aligned}\tilde{Y}_j(t) =& \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\int_{-\infty}^{t} e^{\frac{\alpha}{2}s} d\tilde{B}_j(s),\\ \tilde{X}^{(1*)}(t) =& \sum_{j=1}^{2}\tilde{Y}_j^2(t),\\ \mathcal{V}(s) =& \frac{4L(s)-3\sigma^2}{8}\mathbb{I}_{\{s\leqslant 0\}} + \frac{4L(s)-3\sigma^2}{8\sqrt{X^{(2*)}(s)}}\mathbb{I}_{\{s>0\}}.\end{aligned}$$

Further, let $\{B_s^{(2*)}, s \geqslant 0\}$ and $\{\bar{B}_s^{(2*)}, s \geqslant 0\}$ be two independent Brownian motions and $\tilde{B}_s^{(2*)}$ be a bilateral Brownian motion with $\tilde{B}_s^{(2*)} = B_s^{(2*)}\mathbb{I}_{\{s>0\}} + \bar{B}_{-s}^{(2*)}\mathbb{I}_{\{s\leqslant 0\}}$. Let

$$\tilde{Y}_t = e^{-\frac{\alpha}{2}t}\int_{-\infty}^{t} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds + \frac{\sigma}{2}e^{-\frac{\alpha}{2}t}\int_{-\infty}^{t} e^{\frac{\alpha}{2}s}d\tilde{B}_s^{(2*)}, \tag{3.3.3}$$

and $\tilde{X}^{(2*)}(t) = \tilde{Y}_t^2$. Define $\tilde{X}(t) = \tilde{X}^{(1*)}(t) + \tilde{X}^{(2*)}(t)$. The following proposition shows that the random sequence $\{\tilde{X}(t+k-1), 0 < t \leqslant 1\}_{k=1}^{\infty}$ is stationary and ergodic.

Proposition 3.3.3. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then,* $\left\{\tilde{X}(t+k-1), 0 < t \leqslant 1\right\}_{k=1}^{\infty}$ *is stationary and ergodic.*

The proof of this result is given in the Appendix B.2. By using this proposition, the asymptotic normality of the estimators is given. As intermediate result, the following proposition states the relation between the auxiliary process and the solution of the SDE (3.2.1).

Proposition 3.3.4. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then, (1)* $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{a.s.} 0$, *, (2)* $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{L^1} 0$.

The proof of this proposition is given in Appendix B.3.

Proposition 3.3.5. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then,*

$$\sup_{t\geqslant 0}\mathbb{E}\left[\left|\tilde{X}(t)\right|\right] < \infty.$$

The proof of this proposition is given in Appendix B.3. To use the stationarity and ergodicity of the auxiliary process, the following proposition is needed to hold.

Proposition 3.3.6. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then,*

$$0 < \sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right] < \infty.$$

The proof of this proposition is given in Appendix B.3.

Proposition 3.3.7. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then,* $\tilde{X}^{-1}(t) - X^{-1}(t) \xrightarrow[t\to\infty]{a.s.} 0$.

Proof. The proof follows from Proposition B.3. □

Proposition 3.3.6 guarantees that the first negative moment of the process $\left\{\tilde{X}(t+k-1), 0 < t \leqslant 1\right\}_{k\in\mathbb{N}_+}$ exists. Further, since the function $y = 1/x, x > 0$ is a measurable function, by *Birkhoff Ergodic Theorem*, $\lim_{L\to\infty}\frac{1}{L}\sum_{k=1}^{L} 1/\tilde{X}(t+k-1) = \mathbb{E}\left[1/\tilde{X}(t)\right]$ a.s. From Proposition 3.3.6, for $s \geqslant 0$, $\sup_{s\geqslant 0}\mathbb{E}\left[1/\tilde{X}(s))\right] < \infty$. Further, due to the convexity of the function $y = \dfrac{1}{x}$ on the interval $(0, +\infty)$, by Jensen's inequality,

$$\mathbb{E}\left[1/\tilde{X}(t)\right] \geqslant 1/\mathbb{E}\left[\tilde{X}(t)\right].$$

The equality "=" holds if and only if $\tilde{X}(t)$ is a constant with probability 1. Since $\tilde{X}(t)$ is a continuous random variable for all $t \geqslant 0$, $\mathbb{E}\left[1/\tilde{X}(t)\right] > 1/\mathbb{E}\left[\tilde{X}(t)\right] > 0$. Finally,

$$0 < 1/\mathbb{E}\left[\tilde{X}(t)\right] < \mathbb{E}\left[1/\tilde{X}(t)\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right] < \infty, \qquad \text{for all } t \geqslant 0. \tag{3.3.4}$$

The relation (3.3.4) plays a crucial role in proving the asymptotic normality of the estimators.

3.3.3 Joint asymptotic normality of the UMLE and RMLE

In this subsection, the joint asymptotic normality of the UMLE and RMLE is derived. The derived limiting distributions play an important role in constructing shrinkage estimators and studying their asymptotic relative efficiency. The established limiting distribution is also important in constructing a test for the hypothesis testing problem (2.2.3). To this end, the asymptotic behavior of the positive definite matrix $\frac{1}{T}Q_{[0,T]}$ and the column vector $\frac{1}{\sqrt{T}}W_{[0,T]}$ is studied. To simplify some mathematical expressions, let

$$\Sigma = \begin{bmatrix} \int_0^1 \varphi^\top(t)\varphi(t)\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]dt & -\int_0^1 \varphi^\top(t)dt \\ -\int_0^1 \varphi(t)dt & \int_0^1 \mathbb{E}\left[\tilde{X}(t)\right]dt \end{bmatrix}_{(p+1)\times(p+1)}. \tag{3.3.5}$$

The following proposition shows that the matrix Σ is invertible.

Proposition 3.3.8. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then, Σ is a positive definite matrix.*

The proof of this proposition is given in Appendix B.3. This result is used in deriving the asymptotic normality of the UMLE. As intermediate step, let us first note that

$$\sqrt{T}\left(\hat{\theta}_T - \theta\right) = \sigma T Q^{-1}_{[0,T]}\frac{1}{\sqrt{T}}W_{[0,T]}. \tag{3.3.6}$$

Thus, the asymptotic behavior of $\sqrt{T}(\hat{\theta}_T - \theta)$ relies on the matrix $TQ^{-1}_{[0,T]}$ and the column vector $\frac{1}{\sqrt{T}}W_{[0,T]}$ as T tends to infinity. In the two following propositions, the convergence of these quantities is studied.

Proposition 3.3.9. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then*

$$\frac{1}{T}Q_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \Sigma \text{ and } TQ^{-1}_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \Sigma^{-1}.$$

The proof of this proposition is given in Appendix B.3. The following proposition investigates the L^2-boundedness of $\frac{1}{\sqrt{T}}W_{[0,T]}$, which is useful in proving the convergence of $\hat{\theta}_T$.

Proposition 3.3.10. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then,*
(1) $\dfrac{1}{\sqrt{T}}W_{[0,T]}$ *is* L^2*-bounded; (2)* $\dfrac{1}{T}W_{[0,T]} \xrightarrow[T\to\infty]{a.s.} 0.$

The proof of this proposition is given in Appendix B.3. The following proposition is used to establish the limiting distribution of $\hat{\theta}_T$. For $0 \leqslant s \leqslant 1$, let

$$W^{(T)}(s) = \left(\frac{1}{\sqrt{T}}\int_0^{sT}\frac{\varphi_1(t)}{\sqrt{X(t)}}dB_t, \ldots, \frac{1}{\sqrt{T}}\int_0^{sT}\frac{\varphi_p(t)}{\sqrt{X(t)}}dB_t, -\frac{1}{\sqrt{T}}\int_0^{sT}\frac{X(t)}{\sqrt{X(t)}}dB_t\right)^\top$$

and

$$\widetilde{W}^{(T)}(s) = \left(\frac{1}{\sqrt{T}}\int_0^{sT}\frac{\varphi_1(t)}{\sqrt{\tilde{X}(t)}}dB_t, \ldots, \frac{1}{\sqrt{T}}\int_0^{sT}\frac{\varphi_p(t)}{\sqrt{\tilde{X}(t)}}dB_t, -\frac{1}{\sqrt{T}}\int_0^{sT}\frac{\tilde{X}(t)}{\sqrt{\tilde{X}(t)}}dB_t\right)^\top.$$

The stationary and ergodic property of the process $\{\tilde{X}(t+k-1), 0 < t \leqslant 1\}_{k\in\mathbb{N}_+}$ helps derive the limiting distribution of $W^{(T)}(s)$. The following lemma shows that the difference between $W^{(T)}(s)$ and $\widetilde{W}^{(T)}(s)$ converges in probability to 0. The derived result gives also the weak convergence of the functional random process $\widetilde{W}^{(T)}(s)$. To this end, let $C_{p+1}[0,1]$ be the space of continuous $p+1$ dimensional functions vector on the closed interval $[0,1]$.

Lemma 3.3.1. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then,*
(1) $W^{(T)}(s) - \widetilde{W}^{(T)}(s) \xrightarrow[T\to\infty]{\mathrm{P}} 0$*;*
(2) *the* $p+1$*-dimensional functional process* $\left\{\widetilde{W}^{(T)}(s), T > 0\right\}_{0\leqslant s\leqslant 1}$ *is tight on the space* $C_{p+1}[0,1]$ *with the uniform topology;*
(3) $\widetilde{W}^{(T)}(s) \xrightarrow[T\to\infty]{\mathrm{D}} \widetilde{W}^*(s)$*, where* $\widetilde{W}^*(s)$ *is a* $p+1$*-dimensional Gaussian process with mean 0 and* $\mathrm{Cov}\left(\widetilde{W}^*(s), \widetilde{W}^*(u)\right) = (s \wedge u)\Sigma$*, for* $0 \leqslant s, u \leqslant 1$.

The proof of this proposition is given in Appendix B.3. It should be noticed that, if $s = 1$, Part **(3)** of Lemma 3.3.1 constitutes a kind of central limit theorem. More precisely,

$$\frac{1}{\sqrt{T}} W_{[0,T]} = W^{(T)}(1) \xrightarrow[T\to\infty]{\mathrm{D}} \widetilde{W}^{*}(1) \sim \mathcal{N}_{p+1}(0, \Sigma). \tag{3.3.7}$$

Proposition 3.3.9- Proposition 3.3.10 and Lemma 3.3.1 imply the asymptotic normality of the UMLE. Thereafter, let $\rho_T = \sqrt{T}(\hat{\theta}_T - \theta)$.

Proposition 3.3.11. *Suppose that Assumption 2.1-2.2, and Assumption 3.1-3.2 hold. Then, $\hat{\theta}_T$ is strongly consistent, and $\hat{\theta}_T$ is asymptotically normal, i.e. $\rho_T \xrightarrow[T\to\infty]{\mathrm{D}} \rho \sim \mathcal{N}_{p+1}(0, \sigma^2\Sigma^{-1})$.*

The proof of this proposition is given in Appendix B.3. Let $\{\mathrm{P}^{(\theta)}_{X^T}\}$ denote the distribution law of the solution of the GCIR model (3.2.1) under the parameter $\theta \in \Theta$. The following theorem gives the property of locally asymptotically normal (LAN) of the probability measures $\{\mathrm{P}^{(\theta)}_{X^T}\}$.

Theorem 3.3.1. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, then, for $\theta_0 \in \Theta$, and any $\boldsymbol{h} \in \mathbb{R}^{p+1}$, the likelihood ratio $Z_T(h) = L\left(\theta_0 + \frac{1}{\sqrt{T}}\boldsymbol{h}, \theta; X^T\right)$ admits the following representation*

$$Z_T(h) = \exp\left\{\left(\boldsymbol{h}, \Delta_T(\theta_0, X^T)\right) - \frac{1}{2}(\Sigma\boldsymbol{h}, \boldsymbol{h}) + r_T(\theta_0, \boldsymbol{h}, X^T)\right\},$$

where $\Delta_T(\theta_0, X^T) \xrightarrow[T\to\infty]{\mathrm{D}} \mathcal{N}(0, \Sigma)$ and $r_T(\theta_0, \boldsymbol{h}, X^T) \xrightarrow[T\to\infty]{\mathrm{P}_{\theta_0}} 0$.

The proof of this proposition is given in Appendix B.3.

Remark 3.3.2. Theorem 3.3.1 shows that the family of measures $\{\mathrm{P}^{(\theta)}_{X^T}\}$ is LAN at every point $\theta \in \Theta$, with local scale $\frac{1}{\sqrt{T}}$ and matrix Σ.

In the following, the joint asymptotic normality of the UMLE, the RMLE and their difference is derived. To this end, the set of local alternative restrictions (2.3.7) is under consideration. Below, the asymptotic normality of $\sqrt{T}(\hat{\theta}_T - \theta)$, $\sqrt{T}(\tilde{\theta}_T - \theta)$ and $\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T)$ is derived. First, note that

$$\sqrt{T}(\tilde{\theta}_T - \theta) = \sqrt{T}\left(I_{p+1} - G_{[0,T]}M\right)\hat{\theta}_T + \sqrt{T}\left(G_{[0,T]}r - \theta\right),$$

and then

$$\varrho_T = \sqrt{T}(\tilde{\theta}_T - \theta) = \left(I_{p+1} - G_{[0,T]}M\right)\sqrt{T}(\hat{\theta}_T - \theta) - \sqrt{T}G_{[0,T]}\left(M\theta - r\right). \qquad (3.3.8)$$

By Proposition 3.3.9 and continuous mapping theorem,

$$G_{[0,T]} \xrightarrow[T\to\infty]{\text{P}} G^* = \Sigma^{-1}M^\top(M\Sigma^{-1}M^\top)^{-1} \qquad (3.3.9)$$

and $I_{p+1} - G_{[0,T]}M \xrightarrow[T\to\infty]{\text{P}} I_{p+1} - G^*M$. Under the set of local alternatives restriction (2.3.7),

$$\sqrt{T}G_{[0,T]}(M\theta - r) = \sqrt{T}G_{[0,T]}\frac{r_0}{\sqrt{T}} = G_{[0,T]}r_0 \xrightarrow[T\to\infty]{\text{P}} G^*r_0. \qquad (3.3.10)$$

Let $(\rho_T^\top, \varrho_T^\top, \varsigma_T^\top)^\top = \sqrt{T}\left((\hat{\theta}_T - \theta)^\top, (\tilde{\theta}_T - \theta)^\top, (\hat{\theta}_T - \tilde{\theta}_T)^\top\right)^\top$, by connecting Proposition 3.3.11, the set of local alternatives restriction (2.3.7) and the convergence in (3.3.10), the following joint asymptotic normality of the UMLE and RMLE is derived.

Proposition 3.3.12. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold along with the set of local alternative restrictions in (2.3.7), then,* $(\rho_T^\top, \varrho_T^\top, \varsigma_T^\top)^\top \xrightarrow[T\to\infty]{\text{D}} (\rho^\top, \varrho^\top, \varsigma^\top)^\top$, *where*

$$\begin{bmatrix}\rho\\ \varrho\\ \varsigma\end{bmatrix} \sim \mathcal{N}_{3(p+1)}\left[\begin{bmatrix}0\\ -G^*r_0\\ G^*r_0\end{bmatrix}, \sigma^2\begin{bmatrix}\Sigma^{-1} & \Sigma^{-1} - G^*M\Sigma^{-1} & G^*M\Sigma^{-1}\\ \Sigma^{-1} - G^*M\Sigma^{-1} & \Sigma^{-1} - G^*M\Sigma^{-1} & 0\\ G^*M\Sigma^{-1} & 0 & G^*M\Sigma^{-1}\end{bmatrix}\right].$$

The proof of this proposition follows Proposition 2.3.7. From this result, in the next section, an asymptotic test for the testing problem (2.2.3) is constructed. The above result is also used, in the next section, for studying the optimality of the proposed estimators.

3.4 Testing the restriction and shrinkage estimators

3.4.1 Testing the restriction

In this subsection, the hypothesis testing problem in (2.2.3) is handled. Note that, the diffusion parameter σ^2 can be consistently estimated by the discretized version of quadratic variation of the process $\left\{2\sqrt{X}(t)), t \geqslant 0\right\}$. Let $\hat{\sigma}^2$ be such an estimator of the diffusion parameter σ^2. Further, let $\chi^2_q(\Delta)$ be the chi-square random variable with q degrees of freedom and non-centrality parameter Δ. In particular, if $\Delta = 0$, $\chi^2_q(\Delta)$ is a (central) chi-square random variable, with q degrees of freedom. Let $\Delta = \frac{1}{\sigma^2} r_0^\top (M\Sigma^{-1}M^\top)^{-1} r_0$, with r_0 given in (2.3.7),

$$\hat{\Gamma} = \frac{1}{\hat{\sigma}^2} M^\top \left(MTQ^{-1}M^\top\right)^{-1} M, \qquad \Gamma = \frac{1}{\sigma^2} M^\top \left(M\Sigma^{-1}M^\top\right)^{-1} M, \tag{3.4.1}$$

$$\psi_T = \varsigma_T^\top \hat{\Gamma} \varsigma_T, \qquad \psi = \varsigma^\top \Gamma \varsigma, \quad \psi_0 = \varsigma_0^\top \Gamma \varsigma_0 \tag{3.4.2}$$

where $\varsigma_T = \sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T)$, and $\varsigma_0 \sim \mathcal{N}_{p+1}\left(0, \sigma^2 G^* M\Sigma^{-1}\right)$.

Proposition 3.4.1. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, together with the set of local alternative restrictions in (2.3.7), then, if* $r_0 \neq 0$, $\psi_T \xrightarrow[T\to\infty]{\text{D}} \psi \sim \chi^2_q(\Delta)$. *Moreover, if* $r_0 = 0$, $\psi_T \xrightarrow[T\to\infty]{\text{D}} \psi_0 \sim \chi^2_q$.

The proof of this proposition is given in Appendix B.4. Let $\chi^2_{\alpha;q}$ be the α^{th} quantile of a χ^2_q where $0 < \alpha \leqslant 1$. Proposition 3.4.1 proposes to reject null hypothesis, in (2.2.3), if $\psi_T > \chi^2_{\alpha;q}$ for a given α, i.e. the suggested test is (2.3.11) Proposition 3.4.1 implies the asymptotic power of the proposed test.

Proposition 3.4.2. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, together with the set of local alternative restrictions in (2.3.7), then, the asymptotic power function of the test in (2.3.11) is given by* $\Pi(\Delta) = \mathrm{P}\left(\chi^2_q(\Delta) \geqslant \chi^2_{\alpha;q}\right)$.

The proof of this Proposition follows from Proposition 3.4.1. It should be noted that, under the null hypothesis, the above asymptotic local power is equal to α. Further, if Δ tends to infinity, the above asymptotic local power tends to 1, which implies that the test is consistent.

To highlight the importance of the established result, let us recall that one of our main goals consists in establishing a statistical method which uses efficiently the sample information as well as the prior knowledge. However, the restriction may not be completely known. In (2.3.11), a consistent test which is useful for testing the hypothesized restriction in (2.2.3) is proposed. Further, in the next subsection, SEs are introduced, which preserve good performance regardless of the validity of the restriction in (2.2.3).

3.4.2 Shrinkage estimators

This subsection presents SEs that represent a compromise between the UMLE and the RMLE. As opposite to the Shrinkage Estimators defined in Section 2.5, the shrinkage estimators proposed in this subsection has known dimensions.

$$\hat{\theta}_T^s = \tilde{\theta}_T + \gamma\left(T\left\|\hat{\theta}_T - \tilde{\theta}_T\right\|^2_{\hat{\Gamma}}\right)\left(\hat{\theta}_T - \tilde{\theta}_T\right), \tag{3.4.3}$$

where $\|x\|^2_A = x^\top A x$, x is a column vector, γ is a continuous real-valued function on $(0, +\infty)$ and $\hat{\Gamma}$ is defined in (3.4.1). It is obvious that if $\gamma(x) = 0$, $\hat{\theta}_T^s = \tilde{\theta}_T$, if $\gamma(x) = 1$, $\hat{\theta}_T^s = \hat{\theta}_T$. As an example, let $\gamma(x) = 1 - (q-2)/x$, with $2 < q = \text{rank}(M) < p+1$, the classical shrinkage estimators (SEs)

$$\hat{\theta}_T^{sh} = \tilde{\theta}_T + \left(1 - (q-2)\psi_T^{-1}\right)\left(\hat{\theta}_T - \tilde{\theta}_T\right), \tag{3.4.4}$$

where ψ_T is the test statistic defined in (3.4.2). Moreover, by taking $\gamma(x) = [1 - (q - 2)/x]^+$, $x > 0$, the positive-part shrinkage estimator (PSE) given by

$$\hat{\theta}_T^{sh+} = \tilde{\theta}_T + \left[1 - (q-2)\psi_T^{-1}\right]^+\left(\hat{\theta}_T - \tilde{\theta}_T\right). \tag{3.4.5}$$

The estimator $\hat{\theta}_T^{sh+}$ has the advantage of avoiding a possible over-shrinking problem.

3.5 Asymptotic distributional risk analysis

This section evaluates the performance of the proposed estimators. The Asymptotic Distributional Risk (ADR) which is given in (2.5.4) is used. For a given estimator $\hat{\theta}_0$ for θ, the ADR of $\hat{\theta}_0$ is defined as

$$\mathrm{ADR}\left(\hat{\theta}_0, \theta; \Omega\right) = \mathbb{E}[(\xi^\top \Omega \xi)] \tag{3.5.1}$$

where ξ is a random vector such that $\sqrt{T}(\hat{\theta}_0 - \theta) \xrightarrow[T\to\infty]{\mathrm{D}} \xi$. In order to fix ideas, $\hat{\theta}_0$ represents an estimator such as $\hat{\theta}_T^{sh}$, $\hat{\theta}_T^{sh+}$, $\hat{\theta}_T$ and $\tilde{\theta}_T$. The following propositions give the ADR of the proposed estimators as well as their asymptotic dominance.

Proposition 3.5.1. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, along with the set of local alternatives in (2.3.7), then,*

$$\mathrm{ADR}\left(\hat{\theta}_T, \theta, \Omega\right) = \sigma^2 \mathrm{trace}(\Omega\Sigma^{-1}).$$
$$\mathrm{ADR}\left(\tilde{\theta}_T, \theta, \Omega\right) = \mathrm{ADR}\left(\hat{\theta}_T, \theta, \Omega\right) - \mathrm{trace}\left(\Omega\sigma^2\left(G^* M \Sigma^{-1}\right)\right) + r_0^\top G^{*\top}\Omega G^* r_0.$$

The proof follows directly from Proposition 3.3.12. Further, let $\Lambda = \Sigma^{-1} - G^* M \Sigma^{-1}$. By using Theorem 3.1 in Nkurunziza [2012], the following proposition gives the ADR of the estimators in (3.4.3).

Proposition 3.5.2. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold along with the set of local alternatives in (2.3.7), then,*

$$\mathrm{ADR}\left(\hat{\theta}_T^{s}, \theta, \Omega\right) = \sigma^2 \mathrm{trace}\left(\Omega\Lambda\right) + r_0^\top G^{*\top}\Omega G^* r_0 - 2\mathbb{E}\left[\gamma\left(\chi^2_{q+2}(\Delta)\right)\right] r_0^\top G^{*\top}\Omega G^* r_0$$
$$+\mathbb{E}\left[\gamma^2\left(\chi^2_{q+2}(\Delta)\right)\right] \mathrm{trace}\left(\Omega\sigma^2 G^* M\Sigma^{-1}\right) + \mathbb{E}\left[\gamma^2\left(\chi^2_{q+4}(\Delta)\right)\right] r_0^\top G^{*\top}\Omega G^* r_0.$$

The proof follows the proof of Proposition A.20. Let $\lambda_{\min}, \lambda_{\max}$ be the smallest and largest eigenvalues of the matrix $\left(G^{*\top}\Gamma G^*\right)^{-1} G^{*\top}\Omega G^*$, respectively. Below,the comparison of the ADR between different estimators is presented.

Proposition 3.5.3. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, along with the set of local alternatives in (2.3.7), then*

$$if\ \Delta \leqslant \sigma^2 \text{trace}(\Omega G^* M \Sigma^{-1})/\lambda_{\max},\ then,\ \text{ADR}\left(\tilde{\theta}_T, \theta, \Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T, \theta, \Omega\right);$$

$$if\ \Delta \geqslant \sigma^2 \text{trace}(\Omega G^* M \Sigma^{-1})/\lambda_{\min},\ then,\ \text{ADR}\left(\tilde{\theta}_T, \theta, \Omega\right) \geqslant \text{ADR}\left(\hat{\theta}_T, \theta, \Omega\right).$$

The proof follows the proof of Proposition 2.5.1. The following proposition confirms the efficiency of the SEs compared with the UMLE.

Proposition 3.5.4. *If Assumption 2.1-2.2, and Assumption 3.1-3.2 hold, along with the set of local alternatives in (2.3.7), then,*

$$\text{ADR}\left(\hat{\theta}_T^{sh+}, \theta, \Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T^{sh}, \theta, \Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T, \theta, \Omega\right)$$

for all $\Delta \geqslant 0$, *provided* $\frac{\sigma^2 \text{trace}(\Omega G^* M \Sigma^{-1})}{\lambda_{\max}} \geqslant \frac{q+2}{2}$.

The proof follows Proposition 2.5.2.

3.6 Empirical study and numerical results

To highlight the performance of the proposed method, this section presents the simulation results which show that the proposed method performs very well in small and medium time horizons. Mnte-Carlo simulation along with the Euler-Maruyama discretization approach is used to generate the observations which follow the GCIR process in (3.2.1), In particular, the GCIR process with a trigonometric orthogonal function system
$\{1,\ \sqrt{2}\cos(\omega t),\ \sqrt{2}\sin(\omega t),\ \sqrt{2}\cos(2\omega t),\ \sqrt{2}\sin(2\omega t)\}$, where $\omega = 2\pi$ is generated. Hence, the simulated GCIR process is given by

$$dX(t) = \left(\mu_1 + \mu_2 \sqrt{2}\cos(2\pi t) + \mu_3 \sqrt{2}\sin(2\pi t) + \mu_4 \sqrt{2}\cos(4\pi t) - \alpha X(t)\right) dt + \sigma \sqrt{X(t)} dB_t, \tag{3.6.1}$$

where the pre-assigned parameter $\theta = (\mu_1, \mu_2, \mu_3, \mu_4, \alpha)^\top = (5, 1, 2, 1, 1)^\top$. The stochastic process with $T = 20$, $T = 35$, $T = 50$ and $T = 80$ is used to evaluate the effect of time T. Let $\delta = 0.001$ be the time increment. Five hundred iterations are performed and, for each iteration, the parameters are estimated, the mean and standard error of the estimators are recorded. First, to test whether the periodic function $L(t)$ is a constant leads to the following null hypothesis $M\theta = 0$, for given

$$M = [0_{3\times 1} \;\vdots\; I_3 \;\vdots\; 0_{3\times 1}]. \tag{3.6.2}$$

3.6.1 Parameter estimation

Notice that, letting $G(t, X(t)) = 2\sqrt{X(t)}$, by Itô's lemma,

$$dG(t, X(t)) = \frac{\partial G}{\partial t}(t, X(t))dt + \frac{\partial G}{\partial X(t)}(t, X(t))dX(t) + \frac{1}{2}\frac{\partial^2 G}{\partial X^2(t)}(t, X(t))d\langle X\rangle_t,$$

where $\langle X\rangle_t$ denotes the quadratic variation of the process $X(t)$. Since $dX(t) = (L(t) - \alpha X(t))dt + \sigma\sqrt{X(t)}dB_t$, then

$$\begin{aligned} dG(t, X(t)) =& \frac{1}{\sqrt{X(t)}}dX(t) + \frac{1}{2}\left(-\frac{1}{2}X(t)^{-3/2}\right)d\langle X\rangle_t \\ =& \left(\frac{L(t)}{\sqrt{X(t)}} - \alpha\sqrt{X(t)} - \frac{1}{4}\sigma^2\frac{1}{\sqrt{X(t)}}\right)dt + \sigma dB_t \end{aligned}$$

which implies that $d\left(2\sqrt{X(t)}\right) = \left(\frac{L(t)}{\sqrt{X(t)}} - \alpha\sqrt{X(t)} - \frac{1}{4}\sigma^2\frac{1}{\sqrt{X(t)}}\right)dt + \sigma dB_t$. Then, σ^2 is a diffusion parameter which can be estimated by $\frac{1}{T}\sum_{i=1}^{n}\left(2\sqrt{X(t_i)} - 2\sqrt{X(t_{i-1})}\right)^2$. Estimators and standard deviations are reported in Table 3.1-3.4.

For the given real-valued function $\varphi(t)$ and the pre-assigned parameter θ, let $r = M\theta$ be the restrictions on the parameters. To show the advantage of the RMLE while the restrictions on parameters are correct, in Table 3.5, the *Relative Mean Squared Error* (RMSE) under this null hypothesis $M\theta = r$ are reported. The results in Table 3.5 show that the RMLE has the best performance among all the four types of proposed estimators. This confirms the established theoretical result given in Section 3.5. Fur-

Table 3.1: Mean and standard deviation of estimators (T=80)

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\alpha}$
UMLE	5.0258	0.9969	2.0002	0.9979	1.0054
	(0.5047)	(0.0385)	(0.0313)	(0.0248)	(0.1018)
RMLE	2.2738	0.0000	0.0000	0.0000	0.4520
	(0.0889)	(0.2259e-15)	(0.3020e-15)	(0.1199e-15)	(0.0172)
SEs	5.0255	0.9968	1.9999	0.9978	1.0054
	(0.5046)	(0.0385)	(0.0313)	(0.0248)	(0.1018)
PSE	5.0255	0.9968	1.9999	0.9978	1.0054
	(0.5046)	(0.0385)	(0.0313)	(0.0248)	(0.1018)

Table 3.2: Mean and standard deviation of estimators (T=50)

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\alpha}$
UMLE	5.0354	0.9974	1.9989	0.9995	1.0072
	(0.5223)	(0.0453)	(0.0338)	(0.0308)	(0.1059)
RMLE	2.8350	0.0000	-0.0000	0.0000	0.5633
	(0.1379)	(0.2171e-15)	(0.3093e-15)	(0.1164e-15)	(0.0267)
SEs	5.0350	0.9972	1.9985	0.9993	1.0071
	(0.5223)	(0.0453)	(0.0368)	(0.0308)	(0.1059)
PSE	5.0350	0.9972	1.9985	0.9993	1.0071
	(0.5223)	(0.0453)	(0.0368)	(0.0308)	(0.1059)

Table 3.3: Mean and standard deviation of estimators (T=35)

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\alpha}$
UMLE	5.0588	0.9910	1.9992	0.9996	1.0128
	(0.5822)	(0.0635)	(0.0539)	(0.0492)	(0.1194)
RMLE	4.2861	0.0000	-0.0000	0.0000	0.8547
	(0.3062)	(0.1474e-15)	(0.2762e-15)	(0.1198e-15)	(0.0604)
SEs	5.0584	0.9906	1.9982	0.9991	1.0127
	(0.5821)	(0.0635)	(0.0539)	(0.0491)	(0.1194)
PSE	5.0584	0.9906	1.9982	0.9991	1.0127
	(0.5821)	(0.0635)	(0.0539)	(0.0491)	(0.1194)

Table 3.4: Mean and standard deviation of estimators (T=20)

	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\mu}_3$	$\hat{\mu}_4$	$\hat{\alpha}$
UMLE	5.0665	0.9959	2.0034	1.0012	1.0140
	(0.5962)	(0.0500)	(0.0442)	(0.0383)	(0.1209)
RMLE	3.3503	0.0000	-0.0000	0.0000	0.6664
	(0.2012)	(0.1707e-15)	(0.28069e-15)	(0.1315e-15)	(0.0391)
SEs	5.0660	0.9956	2.0029	1.0009	1.0139
	(0.5961)	(0.0500)	(0.0442)	(0.0383)	(0.1208)
PSE	5.0660	0.9956	2.0029	1.0009	1.0139
	(0.5961)	(0.0500)	(0.0442)	(0.0383)	(0.1208)

Table 3.5: RMSE under $M\theta = (1,2,1)^\top$

	$T = 20$	$T = 35$	$T = 50$	$T = 80$
RMSE-UMLE	1.0000	1.0000	1.0000	1.0000
RMSE-RMLE	1.7624	1.6337	1.6894	1.9154
RMSE-SEs	1.2206	1.1420	1.1542	1.2482
RMSE-PSE	1.2233	1.2091	1.2099	1.2499

ther, the simulations are performed under the set of local alternative restriction. Let $r_0 = 0.5kr$, $k = 1,2,3,4,5,6$. Δ is computed by using $\Delta = \frac{1}{\sigma^2} r_0^\top (M\Sigma^{-1}M^\top)^{-1} r_0$. Let $\Sigma_p = \int_0^1 \varphi^\top(t)\varphi(t)\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right] dt$ and $\Lambda = \int_0^1 \mathbb{E}\left[\tilde{X}(t)\right] dt$. In this subsection, by Proposition 3.3.9, the Riemann summation of the integral $\frac{1}{T}\int_0^T \frac{\varphi^\top(t)\varphi(t)}{X(t)} dt$, $\frac{1}{T}\int_0^T X(t)dt$ is used to approximate the matrix Σ_p and Λ, respectively. Let $\hat{\Sigma}_p = \frac{1}{T}\sum_{i=1}^{n} \frac{\varphi^\top(t_i)\varphi(t_i)}{X(t_i)}\Delta_i$, $\hat{\Lambda} = \frac{1}{T}\sum_{i=1}^{n} X(t_i)\Delta_i$, for $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$ and $\Delta_i = t_i - t_{i-1}$. The estimates are reported in Table 3.1-3.4. From Table 3.1-3.4, it is clear that the estimates get closer to the pre-assigned values and the standard errors get smaller, as T increases. Figure 3.1 gives the histogram of the estimators when $T = 80$. The portray given by Figure 3.1 is consistent with the result given by Proposition 3.3.11. Indeed, the histograms seem quite symmetric with respect to the pre-assigned values, which corroborates the fact that the UMLE is asymptotically normal.

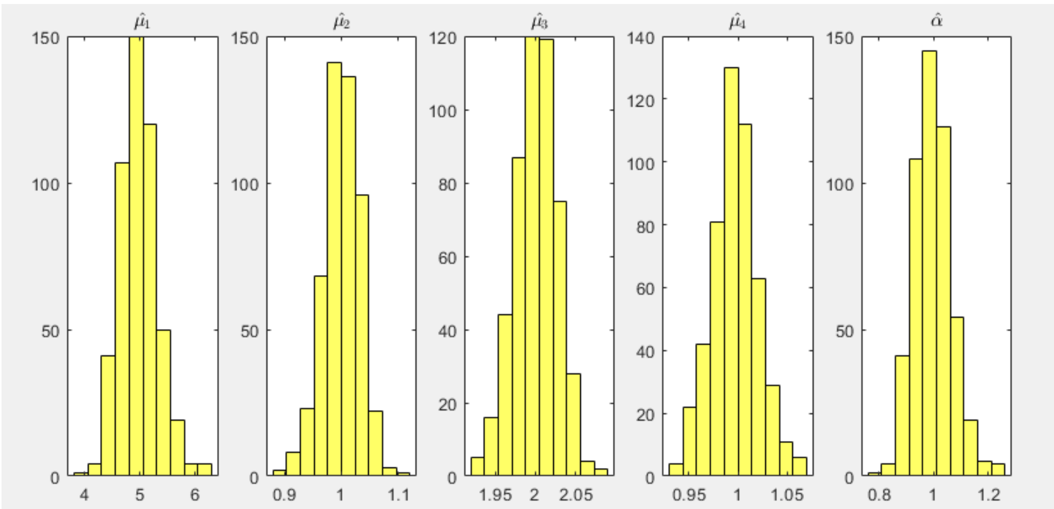

Figure 3.1: The histogram of estimators of GCIR model with $T = 80$

3.6.2 Relative efficiency and empirical power of the test

This subsection sets out with the aim of assessing the performance of the proposed test and estimators versus time T and Δ. To evaluate the performance of the proposed estimators, the relative mean squared efficiency (RMSE) of the proposed estimators is compared. The RMSE of $\hat{\theta}_0$ is defined as $\text{RMSE}(\hat{\theta}_0) = \text{ADR}(\hat{\theta}_T, \theta, \Omega)/\text{ADR}(\hat{\theta}_0, \theta, \Omega)$ where $\hat{\theta}_0$ represents an estimator such as $\tilde{\theta}_T$, $\hat{\theta}_T^{sh}$, $\hat{\theta}_T^{sh+}$ and $\hat{\theta}_T$. For the sake of simplicity, the weighting matrix is taken as $\Omega = I_{p+1}$. The RMSE of each estimators with different time T and different non-centrality parameter Δ is calculated by 500 replications. The obtained RMSE are reported in Figure 3.2-Figure 3.5. These figures are quite revealing in several ways. First, all the figures show that near $\Delta = 0$, RMSE of RMLE is higher than that of the other 3 estimators. This confirms that near the null hypothesis, RMLE is more efficient than the UMLE, SEs and PSE. Second, these figures also indicate that the efficiency of RMLE decreases as Δ is far away from 0. This is consistent with the fact that the RMLE performs worse if the restriction is seriously violated. Furthermore, the figures show that PSE is always more efficient than SEs for all $\Delta \geqslant 0$, which is consistent with Proposition 3.5.4. Meanwhile, RMSE of SEs is decreasing as Δ is far away from 0. However, RMSE of SEs is always higher than RMSE of UMLE for all $\Delta \geqslant 0$.

Another striking observation to emerge from this subsection is the comparison of the variation of the empirical power versus the noncentrality parameter Δ and time T. Figure 3.6 highlights the performance of the proposed test in small and medium time horizon, at the significant level 0.1. Figure 3.6 indicates that the empirical power of the test increases to 1 as Δ increases to infinity and the powers are very close for different time T. This figure confirms the fact that the proposed test is consistent.

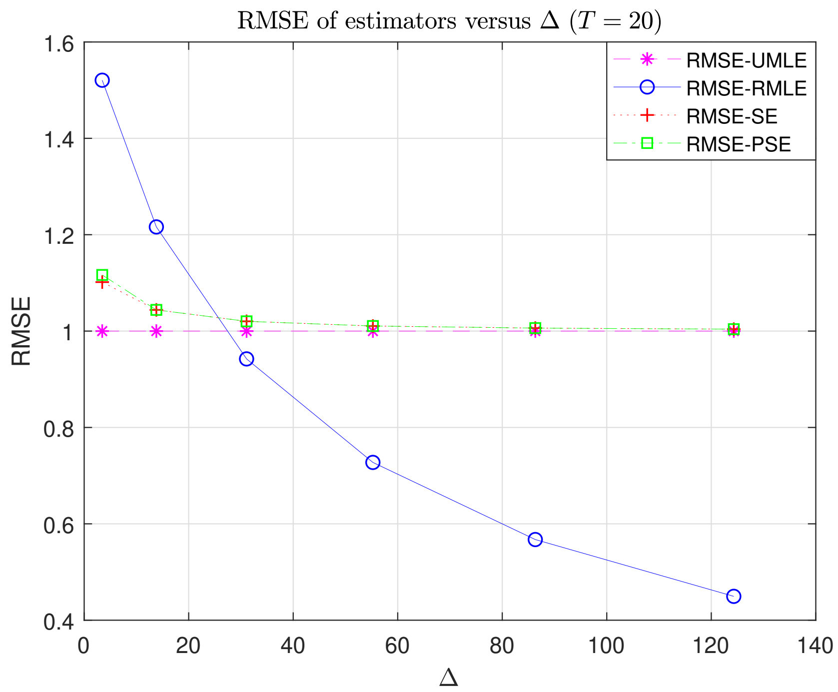

Figure 3.2: Plots of RMSE of UMLE, RMLE, SEs and PSE ($T = 20$)

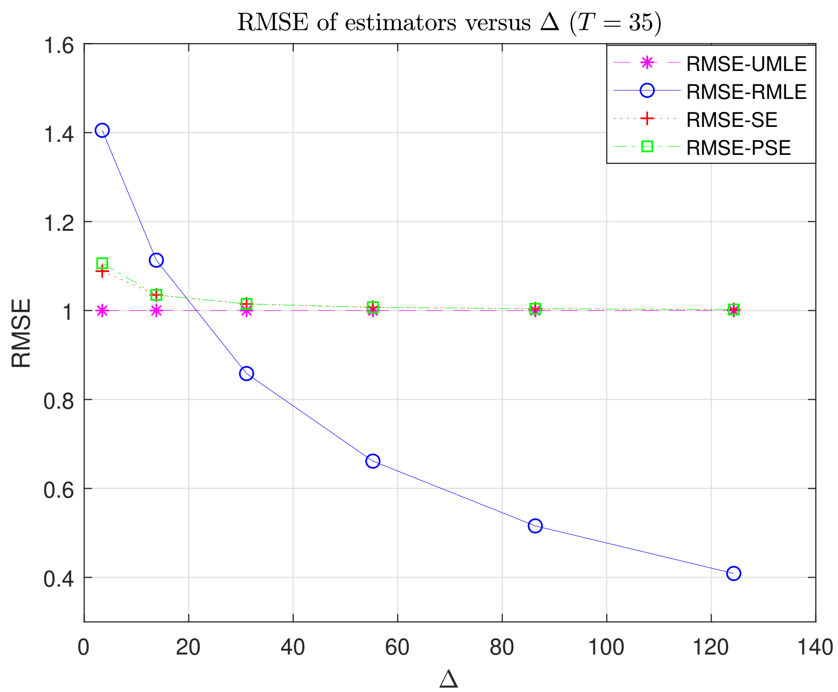

Figure 3.3: Plots of RMSE of UMLE, RMLE, SEs and PSE ($T = 35$)

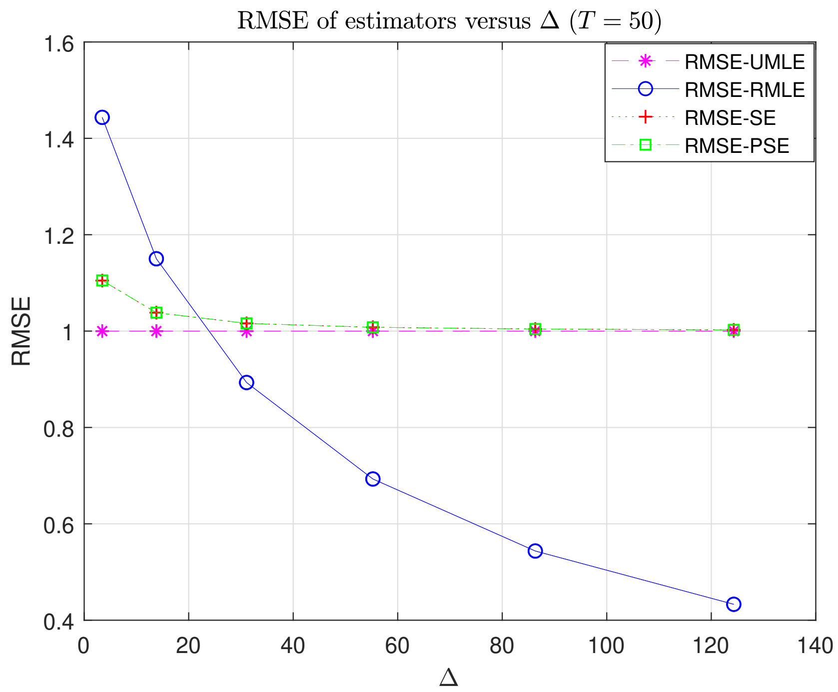

Figure 3.4: Plots of RMSE of UMLE, RMLE, SEs and PSE ($T = 50$)

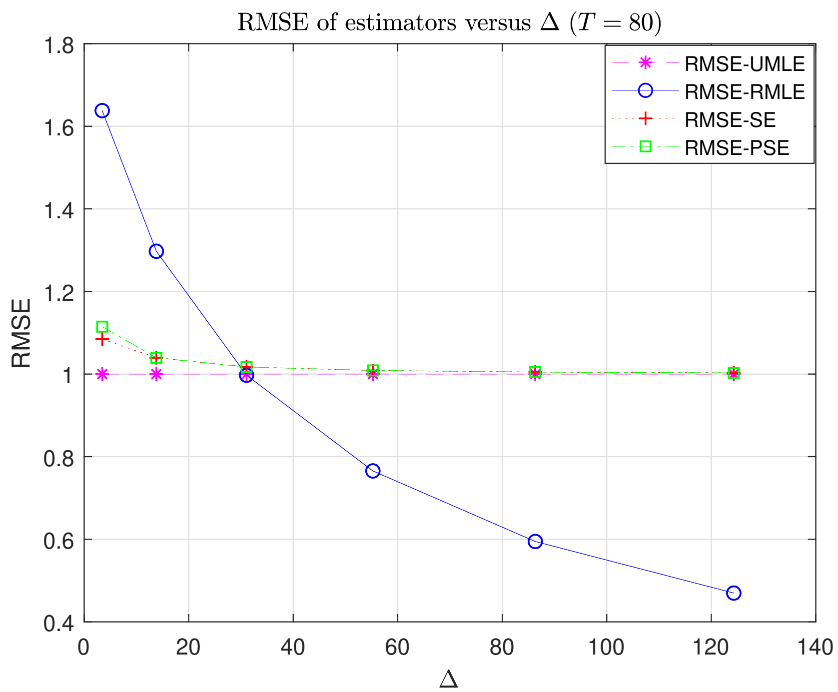

Figure 3.5: Plots of RMSE of UMLE, RMLE, SEs and PSE ($T = 80$)

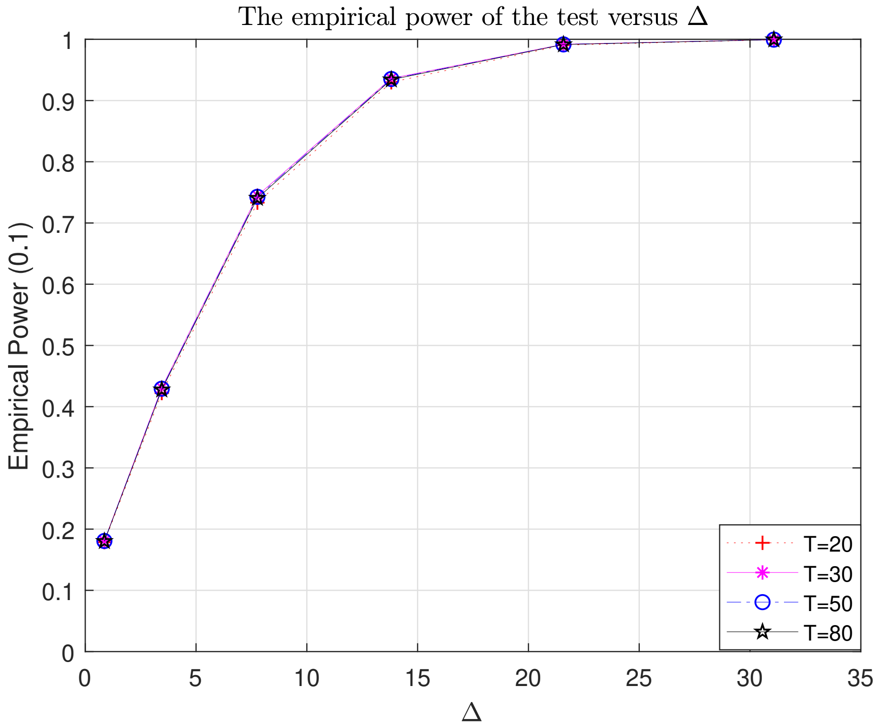

Figure 3.6: Plots of the empirical power of the test versus Δ and T ($\alpha = 0.1$)

3.6.3 Real data analysis

This subsection presents the analysis of real datasets. In particular, Tthe proposed methods are applied to two different datasets. The first dataset under consideration is the historical corn price and the second dataset is the 10-year U.S. treasury bond yield.

3.6.3.1 Historical corn price

This dataset can be found in https://www.macrotrends.net and it represents the corn price recorded daily from 1959 to 2022. The price shown is in U.S. Dollars per bushel. To give a visual description, Figure 3.7 presents the monthly average corn prices from July 01, 1959 to June 27, 2022. As can be seen from the figure, the price of corn gradually increases from January, and by May, the increase reaches its peak. In June the price begins to decline slightly. By July, after the new corn harvest corn prices begin to fall sharply. After September and October, corn prices begin to move higher again. Figure 3.7 shows an obvious periodic pattern of the historical corn prices, which is also in line with the reality of the situation. To apply the method, the observations have been generated by the GCIR process given by

$$dX(t) = (L(t) - \alpha X(t))\, dt + \sigma \sqrt{X(t)} dB_t. \tag{3.6.3}$$

where $L(t) = \mu_1 + \mu_2 \sqrt{2} \cos(2\pi t) + \mu_3 \sqrt{2} \sin(2\pi t) + \mu_4 \sqrt{2} \cos(4\pi t) + \mu_5 \sqrt{2} \sin(4\pi t)$. To apply the proposed method, let $T = 63$, which represents the time span of the data, and $N = 15876$ is the total trading days during 63 years. So the increment of time is $\Delta_N = T/N = 0.004$. For the diffusion parameter σ^2, its estimate is $\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{N} \left(2\sqrt{X(t_i)} - 2\sqrt{X(t_{i-1})}\right)^2 = (0.4061)^2$. The hypothesis testing problem (2.2.3) is

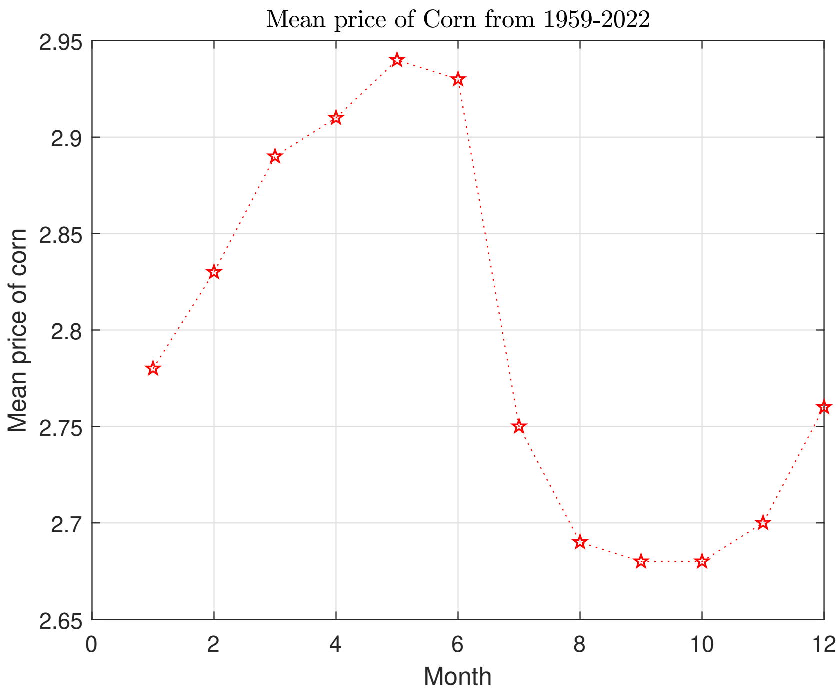

Figure 3.7: Plots of average price of corn from 1959-2022

considered, with $r = (0, 0, 0, 0)^\top$ and M is given as

$$M = [0_{4\times 1} \ \vdots \ I_4 \ \vdots \ 0_{4\times 1}]. \tag{3.6.4}$$

By using the proposed method, the UMLE, RMLE, SEs, and PSE are reported in Table 3.6. The null hypothesis is tested by (2.3.11). The computed test statistic is 27.2932

Table 3.6: UMLE, RMLE, SEs and PSE of historical corn price

Parameter	μ_1	μ_2	μ_3	μ_4	μ_5	α
UMLE	0.2021	-0.2364	-0.0622	-0.1008	-0.0464	0.0467
RMLE	0.1997	-0.0000	-0.0000	-0.0000	0.0000	0.0459
SEs	0.2019	-0.2205	-0.0579	-0.0940	-0.0433	0.0467
PSE	0.2019	-0.2205	-0.0579	-0.0940	-0.0433	0.0467

while the critical value is $\chi^2_{4;0.1} = 7.7794$. Thus, the test statistic value falls into the rejection region at significant level $\alpha = 10\%$. Further, the p-value is 0.00001734. This shows that the null hypothesis should be strongly rejected, that is the ”mean reversion” term is a constant. Figure 3.8 portrays the real historical corn price in daily basis from July 01, 1959 to June 27, 2022 (red) with UMLE. Figure 3.8 also describes the fitted data from model (3.6.3) (blue). What is striking about Figure 3.8is that the predicted data reflects the basic trend of the real bond yield data. *Bootstrap method* on residuals is used to conduct the risk analysis based on 1000 replications. The RMSE is used to compare their relative performance. The values obtained for Bootstrapped RMSE are 0.2816, 1.0169 and 1.0169 for RMLE, SEs and PSE, respectively. The numerical results are in agreement with the theoretical results for which the SEs dominate the UMLE while RMLE performs worse when H_0 is strongly rejected. The figures of the fitted data obtained by using the RMLE, SEs and PSE are similar to Figure 3.8.

3.6.3.2 10-year U.S. treasury bond yield

The 10-year U.S. treasury bond yield is the benchmark used to decide mortgage rates across the U.S. and is the most liquid and widely traded bond in the world. The observations correspond to the total trading days during 11 years (Jan 03 2011- Dec 30 2021). The dataset can be found in http://www.treasury.gov. To give a global visual

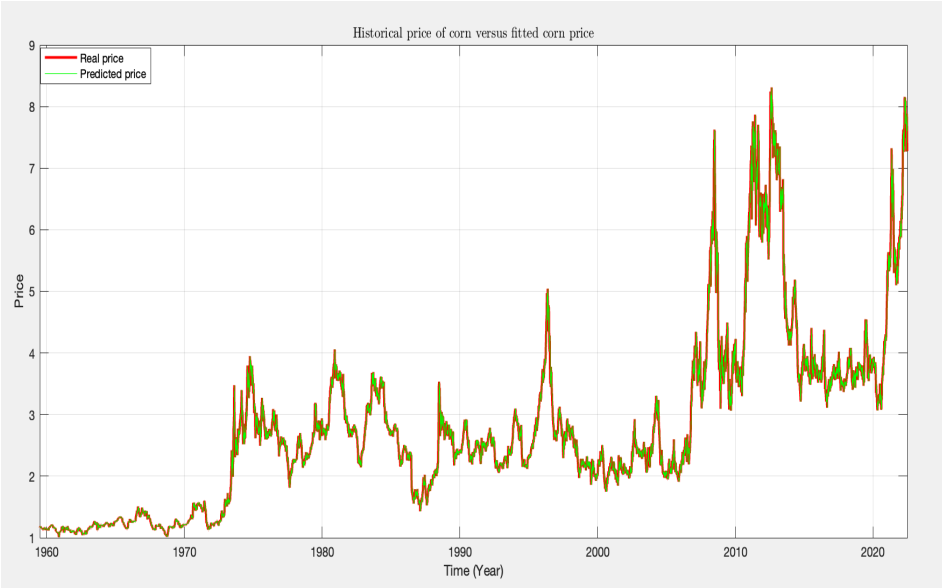

Figure 3.8: Historical price of corn (red) versus fitted corn price (green)

description, Figure 3.9 shows the monthly average price of 10-year U.S. treasury bond yield from 2011 to 2021. From Figure 3.9, it should be noticed that, since the price is significantly affected by policy adjustments, inflation and other factors, periodic pattern in this monthly average curve is not as clear as for the historical corn price. Thus, one can suspect that the restriction in (2.2.3) holds. In other words, it is likely that the mean-reverting level is a constant. This makes it reasonable to consider the hypothesis testing problem (2.2.3), with $r = (0, 0, 0, 0)^\top$ and M given in (3.6.4).

The GCIR model in (3.6.3) is used to apply the proposed methods. The time horizon is taken as $T = 11$, which represents the time span of the data, and $N = 2766$ is the total trading days during 11 years. So the time increment is $\Delta_N = T/N = 0.004$. The diffusion parameter σ^2 is estimated by $\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{N}\left(2\sqrt{X(t_i)} - 2\sqrt{X(t_{i-1})}\right)^2 = (0.5713)^2$. By using the proposed method, the UMLE, RMLE, SEs and PSE are reported in Table 3.7. Further, the obtained test statistic is 7.3471 while the critical value is $\chi^2_{4;0.1} = 7.7794$. Thus, the test statistic does not fall into the rejection region at significant level $\alpha = 0.1$ and the corresponding p-value is 0.11864362. In other words, at 10% significance level,

Figure 3.9: Mean 10-year U.S. Treasury Bond Yield (Jan 03 2011- Dec 30 2021)

Table 3.7: UMLE, RMLE, SEs and PSE of 10-year U.S. treasury bond yield

Parameter	μ_1	μ_2	μ_3	μ_4	μ_5	α
UMLE	1.3208	0.1999	-0.1777	-0.1309	-0.1009	0.6513
RMLE	1.3524	0.0000	0.0000	0.0000	0.0000	0.6664
SEs	1.3331	0.1384	-0.1161	-0.0922	-0.0694	0.6571
PSE	1.3318	0.1409	-0.1208	-0.0928	-0.0712	0.6565

the null hypothesis is failed to reject, that is the ”mean reversion” term is a constant. Figure 3.10 portrays the real yield of 10-year U.S. Treasury bond on a daily basis from January 3, 2011 to December 30, 2021 (Red). Meanwhile, it also describes the fitted data from model (3.6.3) (Blue) with UMLE. What is striking about Figure 3.10 is that the predicted data reflects the basic trend of the real bond yield data. As for the case of historical corn price, the fitted data obtained by the RMLE and SEs give the similar figures. Thus, to save the space of this paper, these figures are omitted. *Bootstrap method* on residuals is used o conduct the risk analysis based on 1000 replications. The RMSE is calculated to compare the relative performance of the proposed estimators. The results are reported in Table 3.8 and these are in agreement with the theoretical results for which the SEs dominate the UMLE while, under H_0, RMLE performs better than the UMLE. This is consistent with the fact that H_0 is not rejected at 10% significance level.

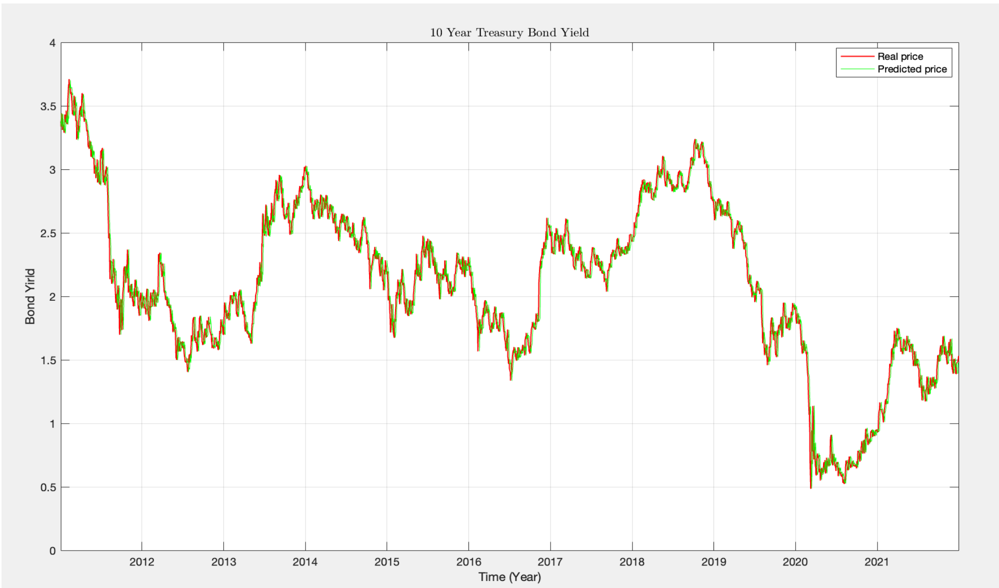

Figure 3.10: 10-year U.S. Treasury Bond Yield (red) versus fitted Treasury Bond Yield (green)

Table 3.8: Bootstrapped RMSE

Estimators	UMLE	RMLE	SEs	PSE
RMSE	1.0000	1.0072	1.0316	1.0371

3.7 Conclusion

This chapter proposed a GCIR process which is suitable for financial datasets which exhibit a periodic mean-reverting level. The proposed stochastic process takes only positive value which is very convenient for many financial datasets. For the drift parameter, different types of estimators including SEs which combine the sample information and the prior information were derived. An asymptotic test was proposed for assessing the prior information given in the form of a linear restriction. The main difficulty of the studied inference problem consists in the fact the GCIR process does not have an explicit solution and such process is not stationary unless it is restricted to the special case where its dimension is a positive integer. To overcome this difficulty, an "*approximate*" auxiliary process which is strictly stationary and ergodic was construced. The difference between the auxiliary process and the solution converges, in L^1 and almost surely, to 0. Based on this result and the stationarity and ergodicity of the auxiliary process, the joint asymptotic normality of the UMLE and RMLE, under the set of local alternative hypotheses was established. UMLE is asymptotically efficient. The derived joint asymptotic normality was used to construct a consistent test for testing the hypothesized restriction. SEs encloses as special cases the UMLE and the RMLE. Further, the joint asymptotic normality was used to derive the asymptotic distributional risk of the proposed estimators which was used to evaluate the relative risk efficiency of the proposed estimators. Specifically, SEs dominate the UMLE and while the RMLE has good performance near the null hypothesis only. Furthermore, the simulations corroborate the theoretical findings. More precisely, the simulation results show that SEs dominate the UMLE. They also show that, near the null hypothesis, the RMLE is better than the

UMLE and SEs. Nevertheless, the performance of RMLE decreases as one moves far away from the restriction. Moreover, PSE is always better than SEs for all $\Delta \geqslant 0$. Finally, the historical corn dataset as well as 10-year U.S. treasury bond yield dataset was used to illustrate the application of the proposed methods. The top novelty of this chapter consists in the fact that the inference were performed under the context that the GCIR process does not have an explicit solution and it is not stationary.

CHAPTER 4

Inference in GCKLS Process

4.1 Introduction

Over about recent three decades, Cox-Ingersoll-Ross (CIR) process has received a lot of attention. Among properties of the CIR, it is a Markovian process and it is views as equilibrium single-term structure model. This process is also known as a square root process that captures the key characteristics of real interest rates. So far, due to its features and capacity along with its mean-reverting property, the CIR process is one of the most commonly used interest rate models in the literature. In particular, Geman [2009] points out that mean-reverting property is found in several applications including commodity and energy price processes. However, despite such a general trend, the assumption of a constant mean-reversion level seems inadequate due to seasonality patterns or a long-term trend of the process. By extending the square root in CIR model into a positive real number, Chan et al. [1992] introduced the so-called Chan, Karolyi, Longstaff and Sanders (CKLS) process to generalize the CIR process. The introduced parameter determines the sensitivity of the variance to the level of the process at some given time point. By using the generalized method of moments (GMM), Chan et al.

[1992] estimated and compared a variety of continuous-time models of the short-term riskless rate. They also examined the performance of the CKLS model. Further, Andersen and Piterbarg [2007] derived the explicit stationary distribution density of the solution and explored the boundary behaviour of a class of stochastic volatility models. The quoted citation Leuwattanachotinan [2011] considered the parameter estimation with CKLS model by applying the GMM and efficient method of moments (EMM) to 3-month UK Repo rates. Another interesting reference is Zíková and Stehlíková [2012] where the authors studied the convergence model of interest rates, which explains how interest rates have changed in connection with the adoption of the Euro currency. In Ying and Hin [2014], all the parameters (drift parameter, diffusion parameter and sensitivity parameter) in CKLS model were considered to be random. A prediction interval for the future value of the interest rate at the next time point when the current value of the interest rate is given. Hu et al. [2015] developed the CKLS model's explicit solution using the Girsanov Theorem and established a link between the CKLS model and CIR process. Cai and Wang [2015] investigated the asymptotic behaviour of a CKLS model randomly disturbed by a small parameter. Under the assumption that the small parameter can be arbitrary small, the central limit theorem and the moderate deviation principle are obtained for the solution of the randomly perturbed CKLS model. Recently, there are some works which studied the parameter estimation of the CKLS model. For instance, under the settings of CKLS model with the sensitivity parameter restricted in the interval $[0, 3/2]$, Sánchez and Gallego [2016] estimated the parameters in two phases when the long-term trend is defined by a continuous deterministic function. Later, Monsalve and Sanchez [2017] studied with the case when the long-term trend of the generalized CKLS model is a deterministic periodic function. In this case, the sensitivity parameter is $0, 1/2$, or 1, and the periodic tendency is represented by the series of Fourier. Wei [2020] considered the parameter estimation problem for discrete observed CKLS model driven by small Lévy noises. The explicit formula of the least squares estimators and the estimation error were derived. Effectiveness of the estimators was also confirmed. Kubilius and Medžiūnas [2020] considered the fractional CKLS model and proved that the trajectories are not necessarily positive if the sensitivity parameter is greater than 1. For solutions to the fractional CKLS model, the almost sure convergence rate of the backward Euler approximation approach was established. Mishura et al. [2022] generalized the estimator in Dehtiar et al. [2021] under the context of CIR process. The strong consistency and asymptotic normality of the maximum likelihood estimator were proved under the context that the sensitivity parameter is restricted to the interval $(0, 1/2)$.

This chapter introduces a generalized process that extends the constant mean reverting to a deterministic and periodic function. The inference problem about the drift parameters is an interesting problem, in the context where some prior information (from outside the sample) may be available. Then, it is attractive to derive a statistical method that combines the prior information and the sample information. Further, to overcome some uncertainty about the restriction, it is interesting to derive a test for testing the hypothesized restriction. The rest of this chapter is structured as follows. Section 4.2 presents the statistical model, some assumptions and preliminary results. Section 4.3 proves that the generalized CKLS model is ergodic, which is the basis of the following estimation and testing work. The parameters of volatility and sensitivity are also estimated. Section 4.4 derives the UMLE and the restricted maximum likelihood estimator (RMLE) of the drift parameter. In this section, the joint asymptotic normality of the UMLE and RMLE, under the set of local alternative restrictions are also established. Section 4.5 derives a test for assessing the hypothesized restriction and a class of shrinkage estimators (SEs) is presented. Section 4.6 establishes the asymptotic distributional risk (ADR) of the proposed estimators and by using the ADR, the asymptotic dominance of these estimators is studied. Section 4.7 presents the empirical study. A data analysis of historical corn prices is performed to demonstrate the application of the proposed method. For more details, we refer to Lyu and Nkurunziza [2023b].

4.2 Statistical model

The CKLS process U_t is the solution to the following Stochastic differential equation (SDE):

$$dU_t = (\beta - \alpha U_t)dt + \sigma U_t^{\delta} dB_t, \quad t \geqslant 0 \tag{4.2.1}$$

where the parameter $\beta > 0$, and the parameter α is the speed of adjustment to the mean reversion level β, σ is the parameter associated with volatility and δ determines the sensitivity of the variance to the level of U_t. In this chapter, the constant parameter β is extended to a real valued function $L(t)$, where t represents the time point. To introduce the proposed model, let $\boldsymbol{\theta} = (\mu_1,\ \mu_2, \ldots,\ \mu_p, \alpha)^{\top} \in \Theta$, where $\Theta \subset \mathbb{R}^{p+1}$ is the parameter space. The following generalized CKLS model is considered

$$dX(t) = S(\theta, t, X(t))dt + \sigma\,(X(t))^{\delta}\, dB_t, \quad X(0) = X_0, \tag{4.2.2}$$

where $S(\theta, t, X(t)) = L(t) - \alpha X(t)$, with $L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t)$. As summarized in Table 4.1, the proposed process in (4.2.2) includes several familiar stochastic processes. Suppose

Table 4.1: Stochastic differential equations with different types of parameters

Model	$L(t)$	β	α	δ	SDE
Merton	β	β	0	0	$dr_t = \beta dt + \sigma dB_t$
Vasicek	β	β	α	0	$dr_t = (\beta - \alpha r_t)dt + \sigma dB_t$
CIR SR	β	β	α	1/2	$dr_t = (\beta - \alpha r_t)dt + \sigma r_t^{1/2} dB_t$
Dothan	β	0	0	1	$dr_t = \sigma r_t dB_t$
GBM	β	0	α	1	$dr_t = -\alpha r_t dt + \sigma r_t dB_t$
Brennan-Schwartz	β	β	α	1	$dr_t = (\beta - \alpha r_t)dt + \sigma r_t dB_t$
CIR VR	β	β	α	3/2	$dr_t = (\beta - \alpha r_t)dt + \sigma r_t^{3/2} dB_t$
CEV	β	0	α	δ	$dr_t = -\alpha r_t dt + \sigma r_t^{\delta} dB_t$
CKLS	β	β	α	δ	$dr_t = (\beta - \alpha r_t)dt + \sigma r_t^{\delta} dB_t$

the target parameter satisfies the linear restriction (2.2.2). As stated in Nkurunziza and

Zhang [2018], the above restriction indicates that there exists some linear relation binding some components of the drift parameter vector. Particularly, when the restriction hold, some improved estimator of θ with high estimation accuracy can be obtained. The restriction (2.2.2) leads to the hypothesis testing problem (2.2.3).

Assumption 4.1. The distribution of the initial value, X_0, of the SDE in (4.2.2) does not depend on the drift parameter θ. X_0 is independent of $\{B_t : t \geqslant 0\}$ and $\mathbb{E}(|X_0|^m) < \infty$, for some $m \geqslant 2$. Further, $X_0 \geqslant U_0$, where U_0 is the initial value of SDE (4.2.1).

In this chapter, the dimension p is assumed to be known, as well as the function $\varphi(t)$. the period v is also supposed to be known and, without loss of generality, $v = 1$. For many financial models, such as option pricing, stochastic volatility, and interest rate models, the positivity is a desired attribute. The following proposition gives a sufficient condition for the process $\{X(t), t \geqslant 0\}$ to be non-negative.

Proposition 4.2.1. *Suppose that Assumption 2.1-2.2,4.1 hold. Then,*

1. *0 is always an attainable boundary for* $0 < \delta < 1/2$.
2. *0 is an unattainable boundary for* $1/2 < \delta$.
3. *∞ is an unattainable boundary for all values of* $0 < \delta$.
4. *0 is an attainable boundary for* $\delta = 1/2$, *if* $2L(t) < \sigma^2$ *and 0 is an unattainable boundary for* $\delta = 1/2$, *if* $2L(t) \geqslant \sigma^2$.

The proof is given in Appendix C.2.

Assumption 4.2. The parameter $1/2 < \delta < 1$ is known, the base function $\varphi(t)$ is analytic, and the function $L(t) \geqslant \beta > 0$, where $\beta > 0$ is the parameter in SDE (4.2.1), and σ is known. Further, $2\alpha - \sigma^2 > 0$.

Remark 4.2.1. In continuous time observations, it is natural to consider that the parameters δ and σ are known. Indeed, these can be obtained explicitly from the observations (see Proposition 4.4.1).

Proposition 4.2.2. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then, the generalized CKLS model in* (4.2.2) *admits a strong and unique non-negative solution,* $X(t)$*, on* $[0, T]$*, for* $0 \leqslant T < \infty$*. Further,* $\{X(t), t \geqslant 0\}$ *is strictly positive almost surely.*

The proof is given in Appendix C.2.

Proposition 4.2.3. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then the solution of the SDE* (4.2.2) *has the form*

$$X(t) = e^{-\alpha t}X_0 + h(t) + Z(t), \quad t \geqslant 0 \tag{4.2.3}$$

where $h(t) = e^{-\alpha t}\int_0^t e^{\alpha s}L(s)ds, Z(t) = \sigma e^{-\alpha t}\int_0^t e^{\alpha s}(X(t))^{\delta}dB_s$.

The proof is given in Appendix C.2.

Proposition 4.2.4. *Suppose that Assumption 2.1-2.2,and Assumption 4.1-4.2 hold. Then,*

$$\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{1-2\delta}\right] < \infty, \quad \sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-2\delta}\right] < \infty, \quad \sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right] < \infty. \tag{4.2.4}$$

The proof is given in Appendix C.2.

Proposition 4.2.5. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then for the process* (4.2.2) *with* $t \geqslant 0$*,*

(1). $\mathbb{E}[X(t)] = e^{-\alpha t}\mathbb{E}[X_0] + \sum_{i=1}^{p}\mu_i\int_0^t e^{-\alpha(t-s)}\varphi_i(s)ds,$ *and* **(2).** $\sup_{t\geqslant 0}\mathbb{E}[X(t)^2] < \infty.$

The proof is given in Appendix C.2. Further, since $1/2 < \delta < 1$, $0 < 2 - 2\delta < 1$, then, Proposition 4.2.5 implies that

$$\sup_{t>0}\mathbb{E}\left[(X(t))^{2-2\delta}\right] < \infty. \tag{4.2.5}$$

Corollary 4.2.1. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then,*

$$\mathrm{P}\left(\int_0^T \left(S(\theta, t, X(t))/(\sigma(X(t))^\delta)\right)^2 dt < \infty\right) = 1, \text{for all } 0 \leqslant T < \infty, \text{ and for all } \theta \in \Theta.$$

The proof is given in Appendix C.2.

4.3 Ergodicity with periodic input

According to [Revuz and Yor, 1999, Chapter III], suppose that there is a process $\{X(t), t \geqslant 0\}$ for which, for any $s < t$, there is a transition probability $\mathbb{P}_{s,t}$ such that

$$P[X(t) \in A|\sigma(X(u), u \leqslant s)] = \mathbb{P}_{s,t}(X(s), A), \quad a.s.$$

Then for the positive function $f : \Omega \mapsto \mathbb{R}_+$, $\mathbb{E}[f(X(t))|\sigma(X(u), u \leqslant s)] = \mathbb{P}_{s,t} f(X(s))$. Particularly, if the process starts at time 0 and given $X_0 = x$, let $\mathbb{P}_{0,t}(x, A)$ as $\mathbb{P}_t(x, A)$. Transition probability $\mathbb{P}_{s,t}$ is written as $\mathbb{P}_{0,t}$. Details are given in Appendix C.

In our case, the input signal is periodic with period 1, and the process $\{X(t), t \geqslant 0\}$ takes values in the interval $(0, +\infty)$ on which the function $x \mapsto x^\delta$ is analytic. Let

$$\mathbb{R}_+^* := \bigcup_{m=1}^{\infty} C_m, \quad with\ C_m := [\frac{1}{m}, m].$$

It is already seen that Assumption C.1 holds. From Assumption 2.2, the grid chain is defined as $\mathbb{X} = (X(k))_{k \in \mathbb{N}_0}$, which is an $(0, +\infty)$-valued time-homogeneous discrete-time Markov process with one-step transition probability $\mathbb{P}_{0,1}$. The path segment chain is also defined: $X = (X(k + s))_{k \in \mathbb{N}_0, 0 < s < 1}$, which is a $(0, 1) \times (0, +\infty)$ valued time-homogeneous continuous-time Markov process.

Proposition 4.3.1. *Suppose that Assumption 2.1-2.2,and Assumption 4.1-4.2 hold. Then, Assumption C.1-C.3 are satisfied for the generalized CKLS model with* $x^* = 1$. *Further,*

1. *The grid chain* $\mathbb{X} = (X(k))_{k \in \mathbb{N}_0}$ *is positive Harris recurrent.*

2. *The path segment chain* $X = (X(k + s))_{k \in \mathbb{N}_0, 0 < s < 1}$ *is positive Harris recurrent.*

The proof of this result is given in Appendix C.2. Then, there exists an unique (up to a multiplicative constant) invariant measure μ for the time-homogeneous grid chain $\mathbb{X} = (X(k))_{k\in\mathbb{N}_0}$. The following theorem gives the strong law of large numbers for the process $\{X(t), t \geqslant 0\}$. Further, let $\{s\}$ be the fractional part of a real number s.

Theorem 4.3.1. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then, for functions* $F : \Omega \mapsto \mathbb{R}$, *which is* $L^1(\mu)$ *bounded,* $\frac{1}{T}\int_0^T F(\{s\}, X(s))ds \xrightarrow[T\to\infty]{a.s.} \left(\int_0^1 \int_{\mathbb{R}} F(s,y)\mu P_{0,s}dy\right)ds$ *for any choice of an initial value* $x \in \Omega$.

Proof. The proof follows from Corollary 2.3 a) [Höpfner et al., 2016, page 531]. □

It is important to note that, the GCKLS model is an extension of the GCIR model discussed in Chapter 3, although Proposition 4.3.1 and Theorem 4.3.1 are established under the assumption that $1/2 < \delta < 1$. Actually, according to the example of CIR-type models in Höpfner et al. [2016], the ergodicity of the GCIR model discussed in Chapter 3 can be derived using the transition probability approach used in this broader class - GCKLS model, provided that the periodic base function $\varphi(t)$ is analytic on the top of satisfying Assumption 3.2 in Chapter 3. Accordingly, the results of Proposition 4.3.1 and Theorem 4.3.1 hold for the special case where the sensitivity parameter $\delta = 1/2$.

4.4 The unrestricted and restricted estimators

This section addresses the maximum likelihood estimator of the target parameters. In particular, the UMLE and the RMLE are derived.

4.4.1 Estimation of the volatility parameters

This subsection gives the estimation of the volatility parameter σ and the sensitivity parameter δ. Let $\langle X\rangle_t = \lim_{n\to\infty}\sum_{k=1}^{2^n}\left(X(\frac{k}{2^n}t) - X(\frac{k-1}{2^n}t)\right)^2$.

Proposition 4.4.1. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then,*

$$\delta = \lim_{h\to 0}\left(\log\left(\frac{\langle X\rangle_{t+h}-\langle X\rangle_t}{\langle X\rangle_{s+h}-\langle X\rangle_s}\right)\Big/\left(2\log\left(X(t)/X(s)\right)\right)\right),$$
$$\sigma^2 = \lim_{h\to 0}\left[\left(\langle X\rangle_{t+h}-\langle X\rangle_t\right)\Big/\left(hX(t)^{2\delta}\right)\right].$$

The proof is similar to the proof of Proposition 5.1-5.3 in Mishura et al. [2022]. For the convenience of the reader, the outline of the proof is given in Appendix C.

4.4.2 The UMLE $\hat{\theta}_T$ and the RMLE $\tilde{\theta}_T$

Let $\mathrm{P}^{(\theta)}_{X^T}$ denote the measure induced by the observable realizations $X^T = \{X(t), t \geqslant 0\}$ on the measurable space $(C_{[0,T]}, \mathcal{B}_{[0,T]})$, where $C_{[0,T]}$ is the space of continuous, real-valued functions on $[0,T]$ and $\mathcal{B}_{[0,T]}$ is the associated Borel σ algebra. Further, let P_B be the measure generated by the Brownian motion on $(C_{[0,T]}, \mathcal{B}_{[0,T]})$. Then, the likelihood function of X^T is given by $\mathcal{L}(\theta, X^T) = (d\mathrm{P}^{(\theta)}_{X^T}/d\mathrm{P}_B)(X^T)$, where $d\mathrm{P}^{(\theta)}_{X^T}/d\mathrm{P}_B$ is the Radon-Nikodym derivative. The UMLE is obtained by maximizing the functional $\boldsymbol{\theta} \longmapsto \mathcal{L}(\theta, X^T)$, i.e. $\hat{\theta}_T = \arg\max_{\theta} \mathcal{L}(\theta, X^T)$. To derive the RMLE, Lagrange multiplier method is used. Let $\tilde{\theta}_T$ be the RMLE, let

$$\boldsymbol{Q}_{[0,T]} = \begin{bmatrix} \displaystyle\int_0^T \frac{\varphi^\top(t)\varphi(t)}{X^{2\delta}(t)}dt & \displaystyle -\int_0^T \varphi^\top(t)X^{1-2\delta}(t)dt \\ \displaystyle -\int_0^T \varphi(t)X^{1-2\delta}(t)dt & \displaystyle\int_0^T X^{2-2\delta}(t)dt \end{bmatrix}_{(p+1)\times(p+1)}, \tag{4.4.1}$$

$$\begin{aligned} R_{[0,T]} &= \left(\int_0^T \frac{\varphi(t)}{X^{2\delta}(t)}dX(t), -\int_0^T \frac{X(t)}{X^{2\delta}(t)}dX(t)\right)^\top, \\ W_{[0,T]} &= \left(\int_0^T \frac{\varphi(t)}{X^{\delta}(t)}dB_t, -\int_0^T \frac{X(t)}{X^{\delta}(t)}dB_t\right)^\top. \end{aligned} \tag{4.4.2}$$

To obtain UMLE $\hat{\theta}_T$, the matrix $\boldsymbol{Q}_{[0,T]}$ needs to be invertible. In the following proposition, it is proven that, under some sufficient conditions, the matrix $\boldsymbol{Q}_{[0,T]}$ is positive definite.

Proposition 4.4.2. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then, if $T \geqslant 1$, $\boldsymbol{Q}_{[0,T]}$ is a positive definite matrix.*

The proof of this result is given in Appendix C.3. Because the optimality of the proposed methods is based on asymptotic results, in the sequel, the condition $T \geqslant 1$ is always supposed to be true. Let $G_{[0,T]} = \boldsymbol{Q}_{[0,T]}^{-1} M^{\top} (M Q_{[0,T]}^{-1} M^{\top})^{-1}$.

Proposition 4.4.3. *Suppose that Assumption 2.1-2.2,and Assumption 4.1-4.2 hold. Then,*

$$\hat{\theta}_T = \boldsymbol{Q}_{[0,T]}^{-1} R_{[0,T]}, \quad \textit{and} \quad \tilde{\theta}_T = \hat{\theta}_T - G_{[0,T]}(M\hat{\theta}_T - r).$$

The proof of this proposition is given in Appendix C.3.

4.4.3 Joint asymptotic normality of the estimators

This subsection derives the joint asymptotic normality of $\hat{\theta}_T$ and $\tilde{\theta}_T$. The limiting distributions play an important role in deriving shrinkage estimators and their asymptotic relative efficiency as well as in deriving an asymptotic test for the hypothesis testing problem (2.2.3). To this end, let us recall that, from Proposition 4.4.3, $\hat{\theta}_T = \boldsymbol{Q}_{[0,T]}^{-1} R_{[0,T]}$. Further, the definition of $R_{[0,T]}$ in (4.4.2) and the SDE (4.2.2) imply that

$$R_{[0,T]} = \boldsymbol{Q}_{[0,T]}\theta + \sigma \int_0^T \frac{(\varphi(t), -X(t))^{\top}}{X^{2\delta}(t)} X^{\delta}(t) dB_t = \boldsymbol{Q}_{[0,T]}\theta + \sigma W_{[0,T]}.$$

Then, $\hat{\theta}_T = \boldsymbol{Q}_{[0,T]}^{-1} \left(\boldsymbol{Q}_{[0,T]}\theta + \sigma W_{[0,T]}\right) = \theta + \sigma \boldsymbol{Q}_{[0,T]}^{-1} W_{[0,T]}$ which implies that

$$\sqrt{T}\left(\hat{\theta}_T - \theta\right) = \sigma T Q_{[0,T]}^{-1} \frac{1}{\sqrt{T}} W_{[0,T]}. \tag{4.4.3}$$

Thus, the asymptotic behavior of $\hat{\theta}_T$ relies on the matrix $T Q_{[0,T]}^{-1}$ and the column vector $\frac{1}{\sqrt{T}} W_{[0,T]}$ as T tends to infinity. To simplify some mathematical expressions, let

$$\boldsymbol{\Sigma} = \begin{bmatrix} \int_0^1 \varphi^{\top}(t)\varphi(t) \int_0^{\infty} y^{-2\delta} \mu P_{0,t} dy dt & -\int_0^1 \varphi^{\top}(t) \int_0^{\infty} y^{1-2\delta} \mu P_{0,t} dy dt \\ -\int_0^1 \varphi(t) \int_0^{\infty} y^{1-2\delta} \mu P_{0,t} dy dt & \int_0^1 \int_0^{\infty} y^{2-2\delta} \mu P_{0,t} dy dt \end{bmatrix}_{(p+1)\times(p+1)}. \tag{4.4.4}$$

Tthe following proposition shows that the matrix $\boldsymbol{\Sigma}$ is invertible. The established result also gives the asymptotic behavior of the matrices $\dfrac{1}{T}\boldsymbol{Q}_{[0,T]}$, as well as $TQ_{[0,T]}^{-1}$.

Proposition 4.4.4. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, then,* **(1).** *the matrix* $\boldsymbol{\Sigma}$ *is a positive definite matrix;* **(2).** $\frac{1}{T}\boldsymbol{Q}_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \boldsymbol{\Sigma}$ *and* $TQ_{[0,T]}^{-1} \xrightarrow[T\to\infty]{a.s.} \boldsymbol{\Sigma}^{-1}$.

Proof. The proof of Part **(1)** is similar to that given for Proposition 4.4.2. The first statement of Part **(2)** follows directly from Theorem 4.3.1. □

The following proposition investigates the L^2-boundedness of $\frac{1}{\sqrt{T}}W_{[0,T]}$, which is useful in proving the convergence of $\hat{\theta}_T$. Further, the established result gives the convergence and asymptotic normality of $\hat{\theta}_T$, which is one of the important results of this chapter. Thereafter, let $\rho_T = \sqrt{T}(\hat{\theta}_T - \theta)$.

Proposition 4.4.5. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold. Then,* **(1)** $\sup\limits_{T\geqslant 0} \mathrm{E}\left(\dfrac{\|W_{[0,T]}\|^2}{T}\right) < \infty$*;* **(2)** $\dfrac{W_{[0,T]}}{T} \xrightarrow[T\to\infty]{a.s.} 0$*;* **(3)** $\frac{1}{\sqrt{T}}W_{[0,T]} \xrightarrow[T\to\infty]{\mathrm{D}} W^* \sim \mathcal{N}_{p+1}(0,\boldsymbol{\Sigma})$*;* **(4)** $\hat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \boldsymbol{\theta}$*; and* $\rho_T \xrightarrow[T\to\infty]{\mathrm{D}} \rho \sim \mathcal{N}_{p+1}(0,\sigma^2\boldsymbol{\Sigma}^{-1})$.

The proof of this proposition is given in Appendix C.3. Let $\{\mathrm{P}_{X^T}^{(\theta)}\}$ denote the distribution law of the solution of (4.2.2) under the parameter $\theta \in \Theta$. The following theorem shows that the locally asymptotic normal (LAN) property of the probability measures $\{\mathrm{P}_{X^T}^{(\theta)}\}$.

Theorem 4.4.1. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold. Then, for* $\theta_0 \in \Theta$*, and any* $\boldsymbol{h} \in \mathbb{R}^{p+1}$*, the likelihood ratio* $Z_T(h) = L\left(\theta_0 + \frac{1}{\sqrt{T}}\boldsymbol{h}, \theta; X^T\right)$ *admits the following representation*

$$Z_T(h) = \exp\left\{\left(\boldsymbol{h}, \Delta_T(\theta_0, X^T)\right) - \frac{1}{2}(\boldsymbol{\Sigma}\boldsymbol{h}, \boldsymbol{h}) + r_T(\theta_0, \boldsymbol{h}, X^T)\right\}$$

where $\Delta_T(\theta_0, X^T) \xrightarrow[T\to\infty]{\mathrm{D}} \mathcal{N}(0,\boldsymbol{\Sigma})$ *and* $r_T(\theta_0, \boldsymbol{h}, X^T) \xrightarrow[T\to\infty]{\mathrm{P}_{\theta_0}} 0$.

The proof of this theorem is given in Appendix C.3.

Remark 4.4.2. Theorem 4.4.1 shows that the family of measures $\{\mathrm{P}_{X^T}^{(\theta)}\}$ is LAN at every point $\theta \in \Theta$, with local scale $\frac{1}{\sqrt{T}}$ and matrix $\mathbf{\Sigma}$.

In the following, the joint asymptotic normality of the UMLE, the RMLE and their difference are presented. Suppose that the set of local alternatives restrictions (2.3.7) holds. Below, the asymptotic normality of $\sqrt{T}(\hat{\theta}_T - \theta)$, $\sqrt{T}(\tilde{\theta}_T - \theta)$ and $\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T)$ is derived. From Proposition 4.4.3,

$$\sqrt{T}(\tilde{\theta}_T - \theta) = \left(\boldsymbol{I}_{p+1} - G_{[0,T]}M\right) \sqrt{T}(\hat{\theta}_T - \theta) - \sqrt{T}G_{[0,T]}\left(\boldsymbol{M\theta} - r\right). \tag{4.4.5}$$

Let $\varrho_T = \sqrt{T}(\tilde{\theta}_T - \theta)$. By Proposition 4.4.4 and continuous mapping theorem,

$$TQ_{[0,T]}^{-1}M^{\top}(MTQ_{[0,T]}^{-1}M^{\top})^{-1} = G_{[0,T]} \xrightarrow[T\to\infty]{\mathrm{P}} G^{*} = \mathbf{\Sigma}^{-1}M^{\top}(M\mathbf{\Sigma}^{-1}M^{\top})^{-1} \tag{4.4.6}$$

and $\boldsymbol{I}_{p+1} - G_{[0,T]}M \xrightarrow[T\to\infty]{\mathrm{P}} \boldsymbol{I}_{p+1} - G^{*}M$. Under the set of local alternatives restriction (2.3.7),

$$\sqrt{T}G_{[0,T]}(\boldsymbol{M\theta} - r) = \sqrt{T}G_{[0,T]}\frac{r_0}{\sqrt{T}} = G_{[0,T]}r_0 \xrightarrow[T\to\infty]{\mathrm{P}} G^{*}r_0. \tag{4.4.7}$$

Let $(\rho_T^{\top}, \varrho_T^{\top}, \varsigma_T^{\top})^{\top} = \sqrt{T}\left((\hat{\theta}_T - \theta)^{\top}, (\tilde{\theta}_T - \theta)^{\top}, (\hat{\theta}_T - \tilde{\theta}_T)^{\top}\right)^{\top}$. By combining Proposition 4.4.5 and relations (2.3.7), (4.4.7), the following joint asymptotic normality of the UMLE and RMLE is established.

Proposition 4.4.6. *Suppose that Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, together with the set of local alternative restrictions in (2.3.7). Then,* $(\rho_T^{\top}, \varrho_T^{\top}, \varsigma_T^{\top})^{\top} \xrightarrow[T\to\infty]{\mathrm{D}} (\rho^{\top}, \varrho^{\top}, \varsigma^{\top})^{\top}$*, where*

$$\begin{bmatrix}\rho\\ \varrho\\ \varsigma\end{bmatrix} \sim \mathcal{N}_{3(p+1)}\left[\begin{bmatrix}0\\ -G^{*}r_0\\ G^{*}r_0\end{bmatrix}, \sigma^2\begin{bmatrix}\mathbf{\Sigma}^{-1} & \mathbf{\Sigma}^{-1} - G^{*}M\mathbf{\Sigma}^{-1} & G^{*}M\mathbf{\Sigma}^{-1}\\ \mathbf{\Sigma}^{-1} - G^{*}M\mathbf{\Sigma}^{-1} & \mathbf{\Sigma}^{-1} - G^{*}M\mathbf{\Sigma}^{-1} & 0\\ G^{*}M\mathbf{\Sigma}^{-1} & 0 & G^{*}M\mathbf{\Sigma}^{-1}\end{bmatrix}\right].$$

The details of the proof is given in Appendix C.3.

4.5 Testing the restriction and shrinkage estimators

4.5.1 Testing the restriction

This subsection moves on to tackle the hypothesis testing problem in (2.2.3). Note that, from Proposition 4.4.1, in the continuous time observations, the sensitivity parameter δ, the diffusion parameter σ^2 can be consistently estimated . Let $\hat{\sigma}^2$ be the estimator of the diffusion parameter σ^2. Further, let $\chi^2_q(\Delta)$ be the chi-square random variable with q degrees of freedom and non-centrality parameter Δ. In particular, if $\Delta = 0$, $\chi^2_q(\Delta)$ is a (central) chi-square random variable, with q degrees of freedom. To perform the hypothesis test, define $\Delta = \frac{1}{\sigma^2} r_0^\top (M\mathbf{\Sigma}^{-1}M^\top)^{-1} r_0$, with r_0 given in (2.3.7). Let

$$\hat{\Gamma} = \frac{1}{\hat{\sigma}^2} M^\top \left(MTQ^{-1}M^\top\right)^{-1} M, \quad \Gamma = \frac{1}{\sigma^2} M^\top \left(M\mathbf{\Sigma}^{-1}M^\top\right)^{-1} M . \psi_T = \varsigma_T^\top \hat{\Gamma} \varsigma_T, \quad \psi = \varsigma^\top \Gamma \varsigma, \tag{4.5.1}$$

where $\psi_0 = \varsigma_0^\top \Gamma \varsigma_0, \varsigma_T = \sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T)$, and $\varsigma_0 \sim \mathcal{N}_{p+1}\left(0, \sigma^2 G^* M \mathbf{\Sigma}^{-1}\right)$.

Proposition 4.5.1. *If Assumption 2.1-2.2,and Assumption 4.1-4.2, together with the set of local alternative restrictions in (2.3.7). Then, if $r_0 \neq 0$, $\psi_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi \sim \chi^2_q(\Delta)$. Moreover, if $r_0 = 0$, $\psi_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi_0 \sim \chi^2_q$.*

The proof of this proposition is given in Appendix C.3. Let $\chi^2_{\alpha;q}$ be the α^{th} quantile of a χ^2_q where $0 < \alpha \leqslant 1$. In this case, the null hypothesis in (2.2.3) is tested by using the rejection region $\psi_T > \chi^2_{\alpha;q}$ for a given α, i.e. the suggested test is given as (2.3.11). From Proposition 4.5.1, below, the asymptotic power of the proposed test is derived.

Proposition 4.5.2. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, together with the set of local alternative restrictions in* (2.3.7). *Then, the asymptotic power function of the test in* (2.3.11) *is given by* $\Pi(\Delta) = \mathrm{P}\left(\chi^2_q(\Delta) \geqslant \chi^2_{\alpha;q}\right)$.

The proof of this proposition follows directly from Proposition 4.5.1. It should be noted that, under the null hypothesis, the above asymptotic local power is equal to α.

Further, if Δ tends to infinity, the above asymptotic local power tends to 1, which implies that the test is consistent. In summary, the analysis procedures and results obtained from UMLE and RMLE are presented. A consistent test is proposed in this section to assess the restriction on drift parameters. Sometimes, the restriction is not completely known. In next subsection, SEs is introduced, which combines both UMLE and RMLE.

4.5.2 Shrinkage estimators

As proposed in Section 2.5, Shrinkage estimators are a compromise between UMLE and RMLE. For the convenience of the reader, the estimators is stated below.

$$\hat{\theta}_T^s = \tilde{\theta}_T + \gamma\left(T\left\|\hat{\theta}_T - \tilde{\theta}_T\right\|_{\hat{\Gamma}}^2\right)\left(\hat{\theta}_T - \tilde{\theta}_T\right), \tag{4.5.2}$$

where $\hat{\Gamma}$ is defined in (4.5.1). Let $\gamma(x) = 1 - (q-2)/x$, with $2 < q = \text{rank}(M) < p+1$, the shrinkage estimators (SEs) is given as

$$\hat{\theta}_T^{sh} = \tilde{\theta}_T + \left(1 - (q-2)\psi_T^{-1}\right)\left(\hat{\theta}_T - \tilde{\theta}_T\right), \tag{4.5.3}$$

where ψ_T is defined in (4.5.1). Moreover, by taking $\gamma(x) = [1-(q-2)/x]^+, x > 0$, the positive-part shrinkage estimator (PSE) given by

$$\hat{\theta}_T^{sh+} = \tilde{\theta}_T + \left[1 - (q-2)\psi_T^{-1}\right]^+\left(\hat{\theta}_T - \tilde{\theta}_T\right). \tag{4.5.4}$$

4.6 Asymptotic distributional risk analysis

This section also evaluates the performance of the UMLE, the RMLE and the SEs by using the Asymptotic Distributional Risk (ADR), based on the *quadratic loss function*. The following propositions give the ADR of the proposed estimators as well as their asymptotic dominance.

Proposition 4.6.1. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, along with the* *set of local alternatives in* (2.3.7). *Then,*

$$\begin{aligned}\mathrm{ADR}\left(\hat{\theta}_T,\theta,\Omega\right) &= \sigma^2\mathrm{trace}(\Omega\mathbf{\Sigma}^{-1});\\ \mathrm{ADR}\left(\tilde{\theta}_T,\theta,\Omega\right) &= \mathrm{ADR}\left(\hat{\theta}_T,\theta,\Omega\right)-\mathrm{trace}\left(\Omega\sigma^2\left(G^*M\mathbf{\Sigma}^{-1}\right)\right)+r_0^\top G^{*\top}\Omega G^* r_0.\end{aligned}$$

The proof of this result is given in Appendix C.3. Further, let $\Lambda=\mathbf{\Sigma}^{-1}-G^*M\mathbf{\Sigma}^{-1}$. By using Theorem 3.1 in Nkurunziza [2012], the following result is derived, which gives the ADR of SEs.

Proposition 4.6.2. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, along with the set of local alternatives in* (2.3.7). *Then,*

$$\begin{aligned}&\mathrm{ADR}\left(\hat{\theta}_T^s,\theta,\Omega\right)=\sigma^2\mathrm{trace}\left(\Omega\Lambda\right)+r_0^\top G^{*\top}\Omega G^* r_0-2\mathbb{E}\left[\gamma\left(\chi^2_{q+2}\left(\Delta\right)\right)\right]r_0^\top G^{*\top}\Omega G^* r_0\\ &+\mathbb{E}\left[\gamma^2\left(\chi^2_{q+2}\left(\Delta\right)\right)\right]\mathrm{trace}\left(\Omega\sigma^2G^*M\mathbf{\Sigma}^{-1}\right)+\mathbb{E}\left[\gamma^2\left(\chi^2_{q+4}\left(\Delta\right)\right)\right]r_0^\top G^{*\top}\Omega G^* r_0.\end{aligned}$$

The proof of this result is given in Appendix C.3. By choosing suitable functions of $\gamma(x)$, $x\geqslant 0$, the following result is derived, which gives the ADR of SEs.

Proposition 4.6.3. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, along with the set of local alternatives in* (2.3.7). *Then,*

$$\begin{aligned}\mathrm{ADR}\left(\hat{\theta}_T^{sh},\theta,\Omega\right) &= \mathrm{ADR}\left(\hat{\theta}_T,\theta,\Omega\right)+(q+2)(q-2)r_0^\top G^{*\top}\Omega G^* r_0\mathbb{E}\left[\chi^{-4}_{q+4}\left(\Delta\right)\right]\\ &\quad-(q-2)\sigma^2\mathrm{trace}\left(\Omega G^*M\mathbf{\Sigma}^{-1}\right)\left(2\mathbb{E}\left[\chi^{-2}_{q+2}\left(\Delta\right)\right]-(q-2)\mathbb{E}\left[\chi^{-4}_{q+2}\left(\Delta\right)\right]\right); \qquad (4.6.1)\\ \mathrm{ADR}\left(\hat{\theta}_T^{sh+},\theta,\Omega\right) &= \mathrm{ADR}\left(\hat{\theta}_T^{sh},\theta,\Omega\right)+2r_0^\top G^{*\top}\Omega G^* r_0\mathbb{E}\left[(1-(q-2)\chi^{-2}_{q+2}\left(\Delta\right))\mathbb{I}_{\{\chi^2_{q+2}(\Delta)<q-2\}}\right]\\ &\quad-\sigma^2\mathrm{trace}\left(\Omega G^*M\mathbf{\Sigma}^{-1}\right)\mathbb{E}\left[(1-(q-2)\chi^{-2}_{q+2}\left(\Delta\right))^2\mathbb{I}_{\{\chi^2_{q+2}(\Delta)<q-2\}}\right]\\ &\quad-r_0^\top G^{*\top}\Omega G^* r_0\mathbb{E}\left[(1-(q-2)\chi^{-2}_{q+4}\left(\Delta\right))\mathbb{I}_{\{\chi^2_{q+4}(\Delta)<q-2\}}\right].\end{aligned}$$

The proof of this proposition is given in Appendix C.3. Let $\lambda_{\min},\lambda_{\max}$ be the smallest and largest eigenvalues of the matrix $\left(G^{*\top}\Gamma G^*\right)^{-1}G^{*\top}\Omega G^*$, respectively. Below, the comparison of the ADR between different estimators is presented.

Proposition 4.6.4. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, along with the set of local alternatives in (2.3.7). If*
$\Delta \leqslant \sigma^2 \text{trace}(\Omega G^* M \boldsymbol{\Sigma}^{-1})/\lambda_{\max}$, *then,* $\text{ADR}\left(\tilde{\theta}_T, \theta, \Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T, \theta, \Omega\right)$ *and if*
$\Delta \geqslant \sigma^2 \text{trace}(\Omega G^* M \boldsymbol{\Sigma}^{-1})/\lambda_{\min}$, *then,* $\text{ADR}\left(\tilde{\theta}_T, \theta, \Omega\right) \geqslant \text{ADR}\left(\hat{\theta}_T, \theta, \Omega\right)$.

The proof of this proposition folows Proposition 2.5.1. The following proposition shows that the PSE has the lowest ADR and it shows that the SEs dominates the UMLE.

Proposition 4.6.5. *If Assumption 2.1-2.2,and Assumption 4.1-4.2 hold, along with the set of local alternatives in* (2.3.7)*, then,*

$$\text{ADR}\left(\hat{\theta}_T^{sh+}, \theta, \Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T^{sh}, \theta, \Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T, \theta, \Omega\right)$$

for all $\Delta \geqslant 0$*, under the condition* $2\sigma^2 \text{trace}(\Omega G^* M \boldsymbol{\Sigma}^{-1}) \geqslant (q+2)\lambda_{\max}$.

The proof of this proposition follows Proposition 2.5.2.

4.7 Numerical evaluation and analysis of real dataset

4.7.1 Simulation results

This subsection contains the simulation results that demonstrate how effectively the suggested approach works across small and medium time periods. To generate the observations which follow the generalized CKLS process in (4.2.2), Monte-Carlo simulation along with the Euler-Maruyama discretization approach is used. In particular, letting $\omega = 2\pi$, the generalized CKLS process is generated with a trigonometric orthogonal function system $\{1, \sqrt{2}\cos(\omega t), \sqrt{2}\sin(\omega t), \sqrt{2}\cos(2\omega t)\}$. Hence, the simulated process is given

$$dX(t) = \left(\mu_1 + \mu_2\sqrt{2}\cos(2\pi t) + \mu_3\sqrt{2}\sin(2\pi t) + \mu_4\sqrt{2}\cos(4\pi t) - \alpha X(t)\right)dt + \sigma X(t)^{\delta} dB_t, \tag{4.7.1}$$

where the pre-assigned parameter $\theta = (\mu_1, \mu_2, \mu_3, \mu_4, \alpha)^\top = (4, 1, 0.5, 1, 1)^\top$. the stochastic process is generated with $T = 20$, $T = 35$, $T = 50$ and $T = 80$ to evaluate the effect of time T, given $\delta = 0.8, \sigma = 0.3$. Let $\Delta = 0.001$ be the time increment. Five hundred iterations are performed and, for each iteration, the parameter, θ, is estimated, the mean and standard error of the estimators are recorded. First, whether the periodic function $L(t)$ is a constant is tested, which leads to the following null hypothesis $\boldsymbol{M\theta} = 0$, for given M a 3×5-matrix with

$$M = \left[\mathbf{0}_{3\times 1} \quad \vdots \quad \boldsymbol{I}_3 \quad \vdots \quad \mathbf{0}_{3\times 1}\right]. \tag{4.7.2}$$

4.7.1.1 Point estimation

This subsection estimates the drift parameters under the condition that $\delta = 0.8$ and $\sigma = 0.3$. The results are shown in Table 4.2. Further, this subsection also investigates the behavior of the estimators of δ and σ. From Proposition 4.4.1, in order to obtain better numerical results, take several distinct points $t_1, \cdots, t_m$, and use the following quantity instead of (C.31),

$$\hat{\delta} = \sum_{i=1}^{m} \left| \log \left(\frac{\langle X \rangle_{t_i+h} - \langle X \rangle_{t_i}}{\langle X \rangle_{s_i+h} - \langle X \rangle_{s_i}} \right) \right| \Big/ \left(2 \sum_{i=1}^{m} \left| \log \left(X(t_i)/X(s_i) \right) \right| \right). \tag{4.7.3}$$

In the following, notice that, letting $G(t, X(t)) = \frac{1}{1-\delta} X(t)^{1-\delta}$, by Itô's lemma,

$$dG(t, X(t)) = \frac{\partial G}{\partial t}(t, X(t))dt + \frac{\partial G}{\partial X(t)}(t, X(t))dX(t) + \frac{1}{2} \frac{\partial^2 G}{\partial X^2(t)}(t, X(t))d\langle X \rangle_t,$$

where $\langle X \rangle_t$ denotes the quadratic variation of the process $X(t)$. This gives

$$\begin{aligned} dG(t, X(t)) &= X^{-\delta}(t)dX(t) - \frac{1}{2}\delta X^{-\delta-1}(t)d\langle X \rangle_t \\ &= \left(X^{-\delta}(t)(L(t) - \alpha X(t)) - \frac{1}{2}\delta\sigma^2 X^{\delta-1}(t) \right) dt + \sigma dB_t \end{aligned}$$

which implies that $d\left(\frac{1}{1-\delta}X^{1-\delta}(t)\right) = \left(X^{-\delta}(t)(L(t) - \alpha X(t)) - \frac{1}{2}\delta\sigma^2 X^{\delta-1}(t)\right) dt + \sigma dB_t$. Then, the diffusion parameter, σ^2, can be estimated by $\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{n} \left(\frac{X^{1-\hat{\delta}}(t_i) - X^{1-\hat{\delta}}(t_{i-1})}{1-\hat{\delta}} \right)^2$. In the simulation study, the number of distinct points $m = 11$ is chosen. The partition step was set to be 0.001 to calculate the quadratic variation in the process. Let the small increment $h = 2^{-6}$. Then, the volatility parameters δ and σ were estimated by using the pairs $s_i = \frac{i}{2m}, t_i = \frac{i+m}{2m}, i = 1, 2, \cdots, m$. The estimates $\hat{\delta}$ and $\hat{\sigma}$ are reported in Table 4.6. The unrestricted estimate of θ is also computed. Table 4.2 shows that the unrestricted estimates of the components of θ get closer to the pre-assigned values and the standard errors get smaller, as T increases. Further, Figure 4.1 gives the histogram of the unrestricted estimators of the components of θ, when $T = 80$. The visual portray given by Figure 4.1 is consistent with the result of Proposition 4.4.5. Indeed, the histograms seem quite symmetric with respect to the pre-assigned values, with mound-shaped curves.

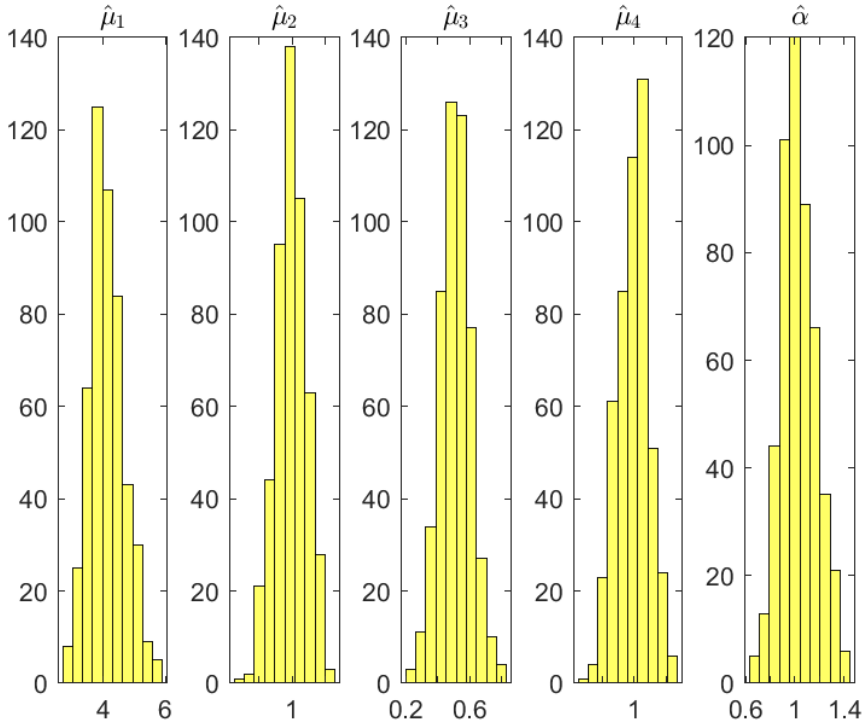

Figure 4.1: The distribution of estimators with $T = 80$

From Figure 4.2, one can see that as T increases, the estimates (using UMLE) of the function $L(t)$ get closer to its true curve. The relative performance of the proposed estimators is also evaluated via the simulation. To this end, the *Relative Mean Squared*

Table 4.2: Mean and standard deviation of estimators of drift parameters (UMLE)

Parameters	μ_1	μ_2	μ_3	μ_4	α
T=20	4.4897	0.9881	0.5132	1.0048	1.1285
	(1.1045)	(0.1957)	(0.1931)	(0.1945)	(0.2954)
T=35	4.2977	0.9988	0.5098	0.9948	1.0781
	(0.8118)	(0.1482)	(0.1514)	(0.1523)	(0.2167)
T=50	4.2215	0.9969	1.5056	0.9946	1.0557
	(0.7143)	(0.1248)	(0.1213)	(0.1216)	(0.1885)
T=80	4.1122	0.9938	0.5109	0.9961	1.312
	(0.5474)	(0.0934)	(0.0973)	(0.0983)	(0.1432)

Table 4.3: Mean and standard deviation of estimators of drift parameters (RMLE)

Parameters	μ_1	μ_2	μ_3	μ_4	α
T=20	4.4390	1.0000	0.5000	1.0000	1.1154
	(1.0565)	(0.2534e-16)	(0.7827e-16)	(0.2166e-16)	(0.2719)
T=35	4.2656	1.0000	0.5000	1.0000	1.0698
	(0.7836)	(0.1217e-16)	(0.2510e-16)	(0.1315e-16)	(0.2089)
T=50	4.1990	1.0000	1.5000	1.0000	1.0499
	(0.6939)	(0.0000)	(0.1648e-16)	(0.8608e-17)	(0.1830)
T=80	4.0860	1.0000	0.5000	1.0000	1.0246
	(0.5301)	(0.4970e-17)	(0.1192e-17)	(0.0000)	(0.1390)

Table 4.4: Mean and standard deviation of estimators of drift parameters (SEs)

Parameters	μ_1	μ_2	μ_3	μ_4	α
T=20	4.4749	0.9946	0.5134	1.0047	1.1247
	(1.0843)	(0.1562)	(0.1516)	(0.1555)	(0.2903)
T=35	4.2854	1.0016	0.5058	0.9998	1.0750
	(0.8009)	(0.1173)	(0.1178)	(0.1239)	(0.2139)
T=50	4.2156	0.9858	1.5038	1.0008	1.0541
	(0.7083)	(0.0953)	(0.0959)	(0.0964)	(0.1868)
T=80	4.1068	0.9944	0.5096	0.9941	1.0299
	(0.5389)	(0.0735)	(0.0760)	(0.0766)	(0.1412)

Table 4.5: Mean and standard deviation of estimators of drift parameters (PSE)

Parameters	μ_1	μ_2	μ_3	μ_4	α
T=20	4.4739	0.9940	0.5105	1.0039	1.1244
	(1.0865)	(0.1458)	(0.1401)	(0.1420)	(0.2908)
T=35	4.2893	1.0006	0.5088	0.9967	1.0760
	(0.8003)	(0.1094)	(0.1109)	(0.1128)	(0.2136)
T=50	4.2137	0.9977	1.5026	0.99712	1.0537
	(0.7071)	(0.0898)	(0.0871)	(0.08793)	(0.1865)
T=80	4.1073	0.9939	0.5093	0.9957	1.0300
	(0.5405)	(0.0676)	(0.0699)	(0.0708)	(0.1415)

Table 4.6: Mean and standard deviation of estimators of parameters δ and σ

Parameters	$T = 20$	$T = 35$	$T = 50$	$T = 80$
δ	0.7632 (0.6035)	0.7608 (0.5375)	0.8087 (0.5789)	0.8117 (0.6059)
σ	0.3157 (0.0016)	0.3169 (0.0013)	0.2971 (9.8165e-04)	0.2958 (7.8919e-04)

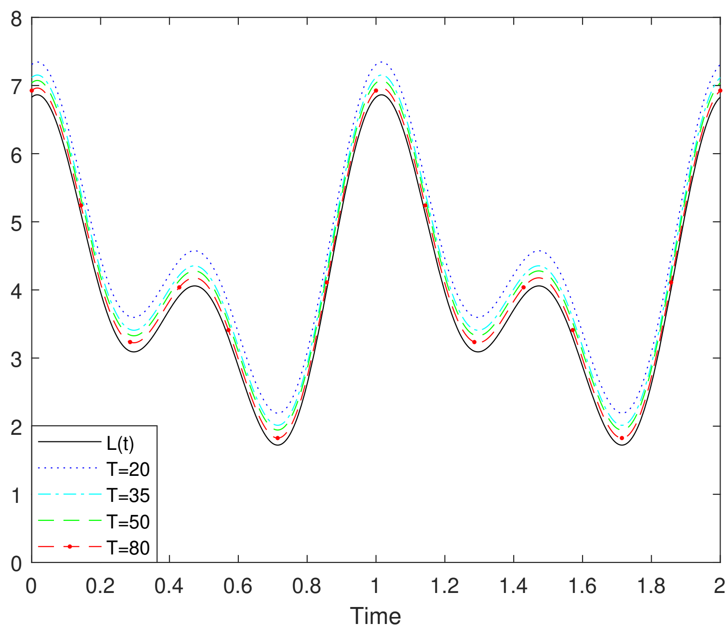

Figure 4.2: The estimation of function $L(t)$ (only show the figures within the interval $[0, 2]$)

Error (RMSE) is calculated as

$$\mathrm{RMSE}(\hat{\theta}_0) = \mathrm{ADR}(\hat{\theta}_T, \theta, \Omega)/\mathrm{ADR}(\hat{\theta}_0, \theta, \Omega) \quad (4.7.4)$$

where $\hat{\theta}_0$ represents an estimator such as $\tilde{\theta}_T$, $\hat{\theta}_T^{sh}$, $\hat{\theta}_T^{sh+}$ and $\hat{\theta}_T$. Thus, to show the advantage of the RMLE while the restriction holds, the RMSE under the null hypothesis is reported in Table 4.7. The results in Table 4.7 show that the RMLE has the best performance among all the four types of proposed estimators. This confirms the established theoretical result given in Proposition 4.6.4. The performance of the UMLE

Table 4.7: RMSE under $\boldsymbol{M\theta} = (1, 1.5, 1)^\top$

	$T = 20$	$T = 35$	$T = 50$	$T = 80$
RMSE-UMLE	1.0000	1.0000	1.0000	1.0000
RMSE-RMLE	1.4300	1.3736	1.3650	1.4218
RMSE-SEs	1.0949	1.1211	1.0924	1.1302
RMSE-PSE	1.1342	1.1390	1.1222	1.1386

and RMLE is assessed, as well as that of SEs and PSEs versus the time horizon, T and the non-centrality parameter Δ. To this end, to generate the process in (4.7.1) under the set of local alternative restrictions, let $r_0 = 0.5kr$, $k = 0, 1, 2, 3, 4, 5$ and let $\Delta = \frac{1}{\sigma^2} r_0^\top (M\boldsymbol{\Sigma}^{-1}M^\top)^{-1} r_0$,

$$\boldsymbol{\Sigma}_p = \int_0^1 \varphi^\top(t)\varphi(t) \int_0^\infty y^{-2\delta}\mu P_{0,t} dydt, \Lambda = \int_0^1 \varphi^\top(t) \int_0^\infty y^{1-2\delta}\mu P_{0,t} dydt,$$

and $\lambda = \int_0^1 \int_0^\infty y^{-2\delta}\mu P_{0,t} dydt$. From Proposition 4.4.4, use the Riemann sum corresponding to the integral $\frac{1}{T}\int_0^T \frac{\varphi^\top(t)\varphi(t)}{X(t)^{2\delta}} dt$, $\frac{1}{T}\int_0^T \varphi^\top(t)X(t)^{1-2\delta} dt$, $\frac{1}{T}\int_0^T X(t)^{2-2\delta} dt$ to approximate the matrix $\boldsymbol{\Sigma}_p$, Λ and λ, respectively. Let $\hat{\boldsymbol{\Sigma}}_p = \frac{1}{T}\sum_{i=1}^n \frac{\varphi^\top(t_i)\varphi(t_i)}{X^{2\delta}(t_i)}\Delta_i$, $\hat{\Lambda} = \frac{1}{T}\sum_{i=1}^n \varphi^\top(t_i)X^{1-2\delta}(t_i)\Delta_i$, $\hat{\lambda} = \frac{1}{T}\sum_{i=1}^n X^{2-2\delta}(t_i)\Delta_i$, for $0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$ and $\Delta_i = t_i - t_{i-1}$. Let $\Omega = \boldsymbol{I}_{p+1}$ be weighting matrix. From 500 replications, the RMSE of the different estimators are obtained. The results are reported in Figures 4.3-4.6. These figures are quite revealing in several ways. First, all the figures show that near $\Delta = 0$, the RMSE of RMLE is the highest, which means that near the null hypothesis, RMLE is more efficient than the

UMLE, SEs, and PSEs. Second, these figures also indicate that the efficiency of RMLE decreases as Δ moves far away from 0. This reflects the fact that the RMLE performs worse if the restriction is seriously violated. Furthermore, such figures show that PSE is always more efficient than SEs for all $\Delta \geqslant 0$, which is consistent with Proposition 4.6.5. Meanwhile, the RMSE of both SEs and PSE are decreasing as Δ is far away from the origin. However, the RMSE of both SEs and PSE are always higher than the RMSE of UMLE for all $\Delta \geqslant 0$.

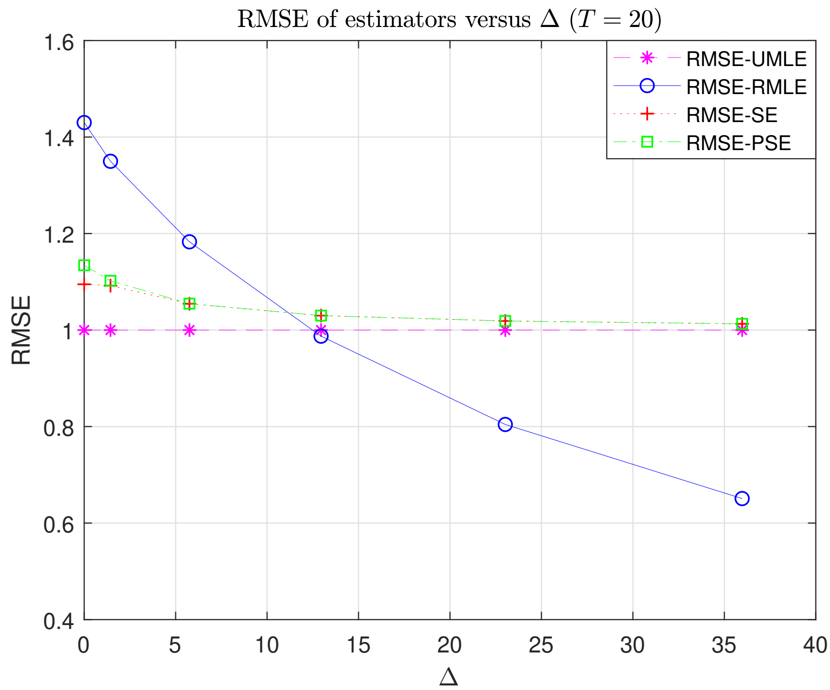

Figure 4.3: The RMSE of the estimator versus Δ ($T = 20$)

4.7.1.2 Empirical power of the test

The performance of the proposed test is also evaluation via simulations. Thus, the variation of the empirical power versus the noncentrality parameter Δ and time T is compared, at the significant level $0.1, 0.05, 0.025$ separately. Figure 4.7-Figure 4.9 show that the empirical power increases to 1 as Δ increases to infinity. These figures also show

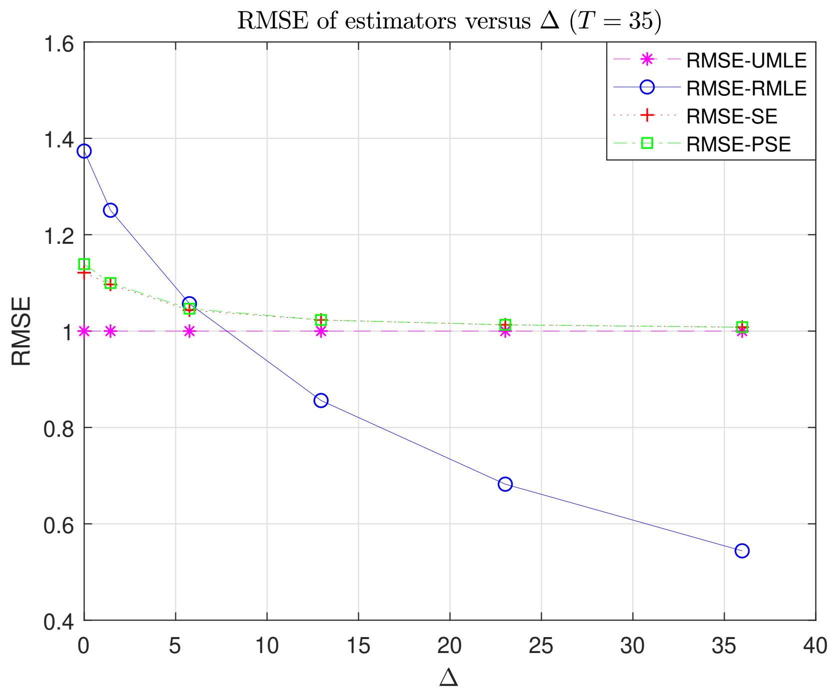

Figure 4.4: The RMSE of the estimator versus Δ ($T = 35$)

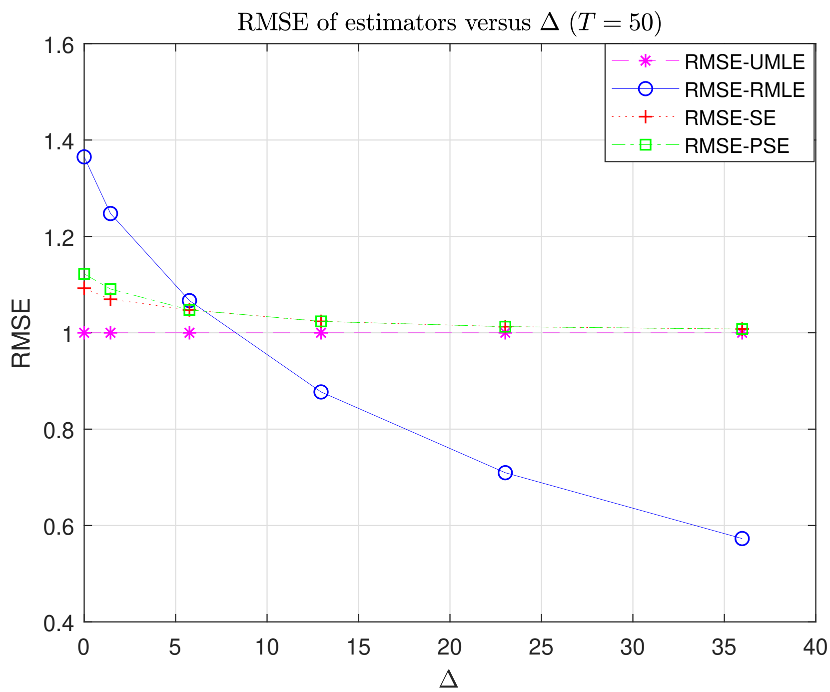

Figure 4.5: The RMSE of the estimator versus Δ ($T = 50$)

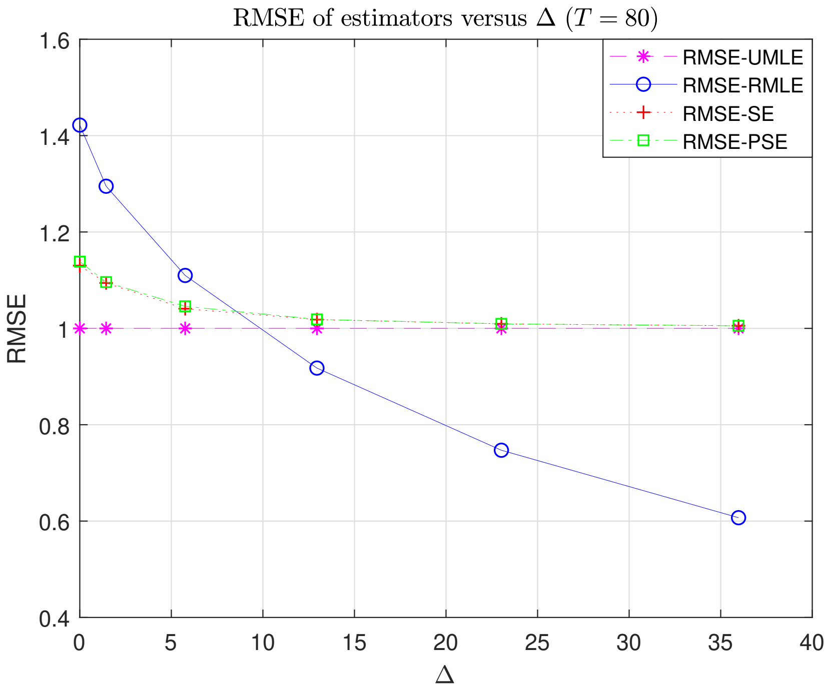

Figure 4.6: The RMSE of the estimator versus Δ ($T = 80$)

that the proposed test is consistent.

4.7.2 Real data analysis

This subsection applies the proposed method to the US Soybean Historical Price recorded daily from 1990 to 2022, which is available at: US-soybeans-historical-data. There are certain seasonal changes in the supply and demand of commodities, i.e., the trend of increase or decrease in the supply or demand of commodities is relatively fixed with the change of seasons, the prices of these commodities are also characterized by seasonal fluctuations, which called the seasonal fluctuation law. In the case of agricultural products, they are usually sown in a specific season of the year, and after growing and maturing, they are harvested in another season. This cycle of growth makes agricultural products have a more obvious seasonal fluctuation law than base metals or chemicals. Figure 4.10 indicates that the price has an obvious periodic property dur-

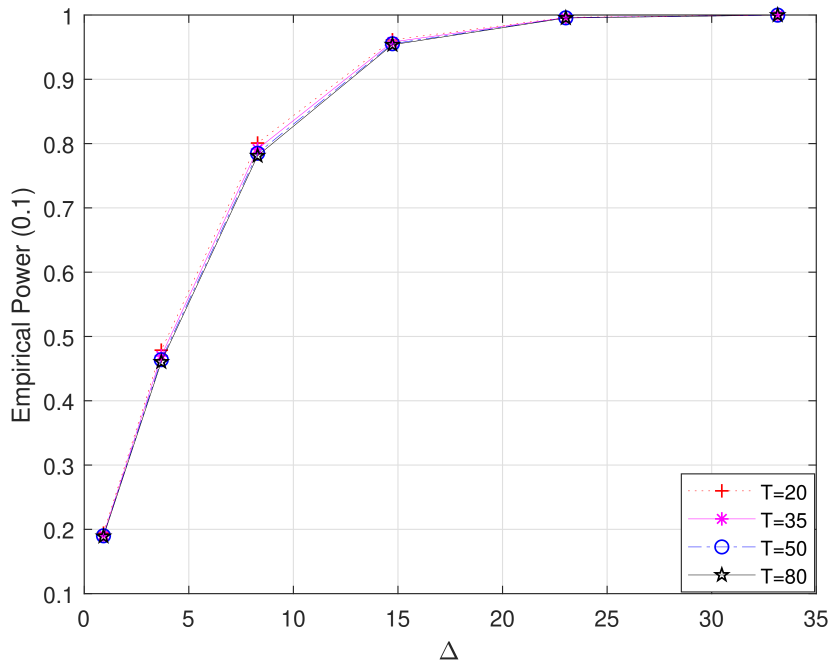

Figure 4.7: Plots of the empirical power of the test versus Δ and T ($\alpha = 0.1$)

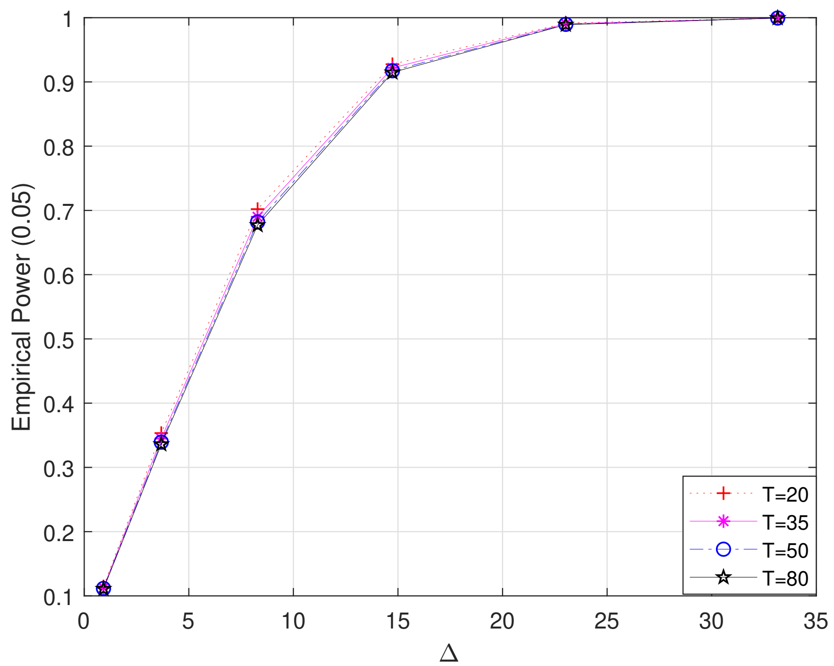

Figure 4.8: Plots of the empirical power of the test versus Δ and T ($\alpha = 0.05$)

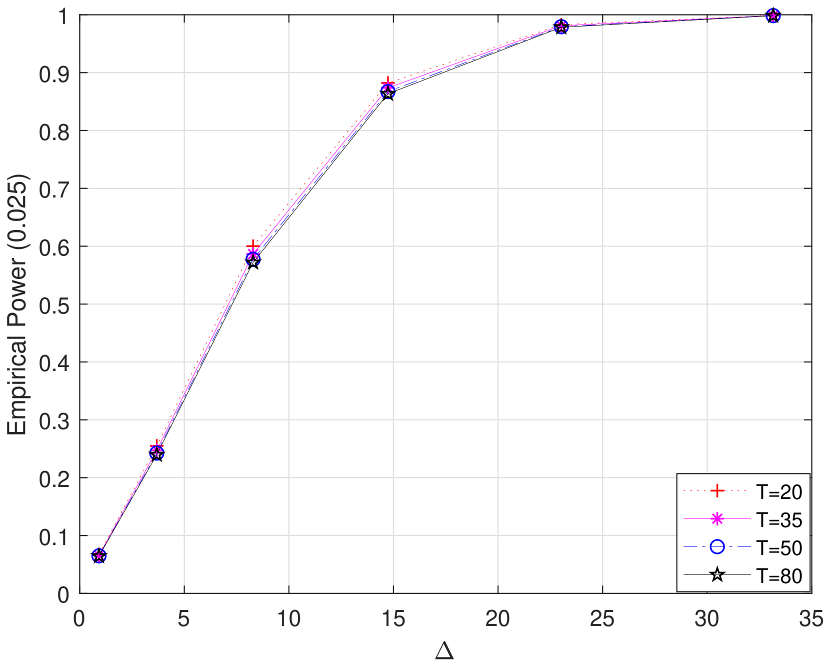

Figure 4.9: Plots of the empirical power of the test versus Δ and T ($\alpha = 0.025$)

ing the year as in the description above. To apply the method, the observations under consideration have been generated by the generalized CKLS process given by

$$dX(t) = (L(t) - \alpha X(t))\, dt + \sigma X^{\delta}(t) dB_t. \tag{4.7.5}$$

where $L(t) = \mu_1 + \mu_2 \sqrt{2} \cos(2\pi t) + \mu_3 \sqrt{2} \sin(2\pi t) + \mu_4 \sqrt{2} \cos(4\pi t) + \mu_5 \sqrt{2} \sin(4\pi t)$. To apply the proposed method, let $T = 33$, which represents the time span of the data, and $N = 8464$ is the total trading days during 23 years. So the increment of time is $\Delta_N = T/N = 0.004$. (4.7.3) implies $\hat{\delta} = 0.6910$ and $\hat{\sigma}^2 = \frac{1}{T} \sum_{i=1}^{N} \left(\frac{1}{1-\hat{\delta}} X^{1-\delta}(t_i) - \frac{1}{1-\hat{\delta}} X^{1-\delta}(t_{i-1}) \right)^2 = (0.0647)^2$. Consider the hypothesis testing problem (2.2.3), with $r = (0, 0, 0, 0)^{\top}$ and M is $4 \times 6-$matrix given as

$$M = \begin{bmatrix} \mathbf{0}_{4\times1} & \vdots & \boldsymbol{I}_4 & \vdots & \mathbf{0}_{4\times1} \end{bmatrix}. \tag{4.7.6}$$

The UMLE, RMLE, SEs and PSE are reported in Table 4.8.

The null hypothesis is testd by using the test in (2.3.11). The computed test statistic

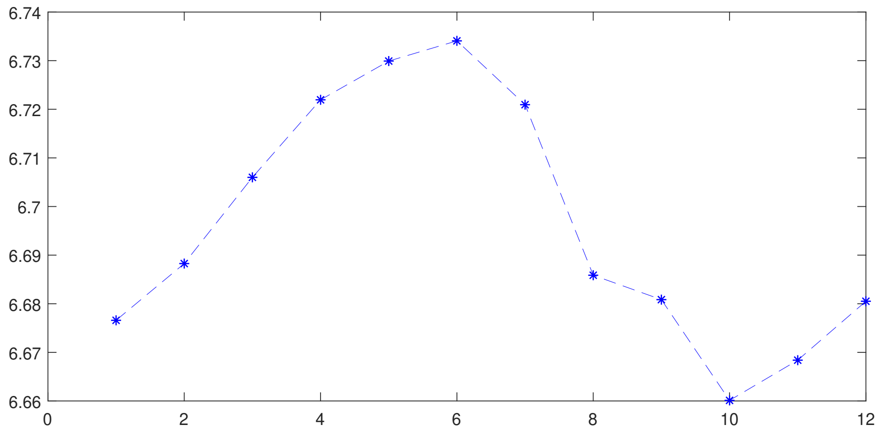

Figure 4.10: The periodic property of US soybean historical log price

Table 4.8: UMLE, RMLE, SEs and PSE of soybean historical data

Parameter	μ_1	μ_2	μ_3	μ_4	μ_5	α
UMLE	1.2773	0.0992	0.0203	0.0395	0.0556	0.1865
RMLE	1.2913	0.0000	0.0000	0.0000	-0.0000	0.1886
SEs	1.2778	0.0959	0.0196	0.0381	0.0537	0.1866
PSE	1.2778	0.0959	0.0196	0.0381	0.0537	0.1866

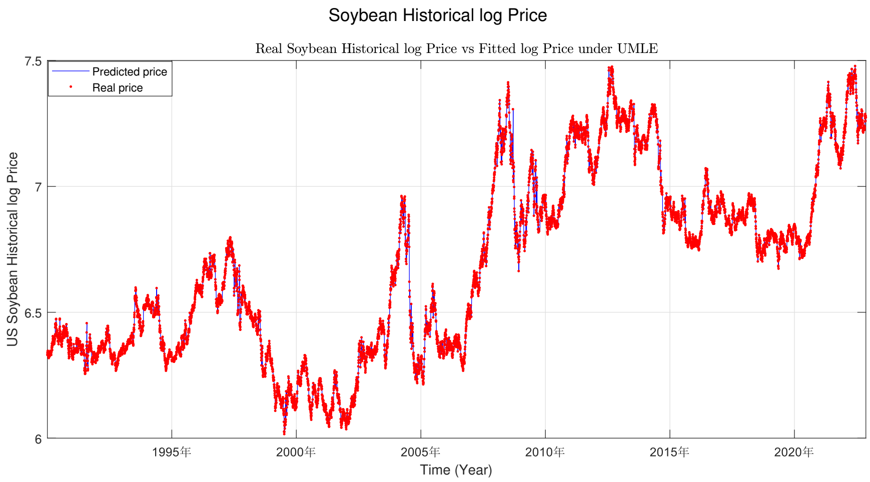

Figure 4.11: Comparison of the real data and fitted value

is 8.5404 while the critical value is $\chi^2_{4;0.1} = 7.7794$. Thus, the test statistic value falls into the rejection region at a significant level $\alpha = 10\%$. Further, the p-value is 0.0737, which is less than $\alpha = 0.1$. This shows that the null hypothesis that the "mean reversion" term is a constant is rejected, at 10% significance level.

4.8 Conclusion

This chapter generalized the CKLS model by extending the constant mean reverting term to a periodic function. The proposed process is more suitable for modeling some financial data, such as the commodity prices, with periodic behaviour. It was proven that the stochastic process takes only positive values under the condition that the sensitive parameter $\delta \in (1/2, 1)$. By using the theory of transition semigroup, it was established that the grid chain corresponding to the generalized CKLS is positive Harris recurrent. Thanks to the established ergodicity, Some inference problems were solved concerning the drift parameters in the context where uncertain prior information is available in the form of a linear restriction on the drift parameter. Three estimators: UMLE, RMLE, and SEs were derived. The joint asymptotic normality of the UMLE and RMLE, under the set of local alternative hypotheses was also established. The derived joint asymptotic normality was used in constructing a consistent test for testing the hypothesized restriction as well as in studying the relative risk efficiency of the proposed estimators. As proved in Section 4.6, SEs dominated the UMLE and while the RMLE had the best performance near the null hypothesis only. Nevertheless, the performance of RMLE decreases as one moves far away from the restriction. Moreover, PSE is always better than SEs for all $\Delta \geqslant 0$. Further, the simulation confirmed the conclusions of the theoretical results. Finally, to illustrate the application of the proposed methods, the historical US soybean dataset, which has obvious seasonality trend, was analyzed.

CHAPTER 5

Summary and Future Research

5.1 Summary

This dissertation introduces three types of stochastic processes that are suitable for positive datasets and that exhibit cyclic mean-reverting level behaviour. In particular, the proposed processes are useful in several financial datasets. The three types of proposed stochastic processes are generalized exponential O-U process, generalized CIR process, and generalized CKLS process.

Firstly, in the context of the generalized exponential O-U process, the parameter estimation and testing of the restrictions are performed under three different cases: no change-point, one known change-point, and one unknown change-point. This process is one of a few cases that admits an explicit solution. In comparison with other existing works about the exponential O-U processes, the proposed GEOU incorporates the seasonality effect and only takes positive values. In addition to that flexibility, the dataset under consideration may be subject to a drastic change. A statistical method is proposed that can be used to validate the seasonality effect or the drastic change as well as other possible relations binding the components of the drift parameter. To this end, the

UMLE and RMLE are derived. In order to derive their joint asymptotic normality as well as the strong consistency of the UMLE, a stationary and ergodic auxiliary process is constructed. The distance between the constructed process and the solution to GEOU model converges to 0 both in mean and almost surely as time tends to infinity. Further, in contrast with the works in recent cited literature, the component of the drift parameter vector is supposed to be unknown. Based on a Schwartz Information Criterion and Euler Approximations, a statistical method is derived to estimate the dimension of the drift parameter and the change-point. An improved test is proposed for testing the dimension and the existence of change-point. The consistency of the test is confirmed in the simulation part.

Secondly, under the context of the proposed generalized CIR process, an inference problem about the drift parameter is studied. For example, there may be an instance where uncertain prior information is available. In particular, SEs combine the sample information and the prior information. An asymptotic test is constructed to assess the prior information given in the form of a linear restriction. The main difficulty of the studied inference problem is twofold: (1) the GCIR process does not have an explicit solution and (2) such processes are not stationary unless they are restricted to the special case where its dimension is a positive integer. An "approximate" auxiliary process that is strictly stationary and ergodic is introduced to overcome this difficulty. Similar to the case of GEOU model, the distance between the auxiliary process and the solution of GCIR process also converges to 0 both in mean and almost surely. By using the stationarity and ergodicity of the auxiliary process, the joint asymptotic normality of the UMLE and RMLE is studied under the set of local alternative hypotheses. UMLE is proven to be asymptotically efficient. The derived joint asymptotic normality is used to construct a consistent test for testing the hypothesized restriction.

Thirdly, GCKLS process is generalized by extending the constant mean reverting term in CKLS model to a periodic function. This is more reasonable to model the commodity prices with obvious seasonality. The stochastic process took only positive values under the condition that the sensitive parameter belongs to the interval (1/2, 1). In the case of GCKLS model, the sensitivity and volatility parameters are supposed to be known. To be more precise, there may be situations where uncertain prior information is available. By using the theory of transition semigroup, it is possible to determine that the grid chain is positive Harris recurrent and that the path segment chain is positive Harris recurrent, from which the strong law of large numbers for the time-inhomogeneous process can be proven. This important finding is used for two purposes: (1) to derive the joint normality of the UMLE and RMLE and (2) to prove the strongly consistency of UMLE. In addition, in all the three different types of stochastic differential equations, the relative performance of the UMLE, RMLE, and the SEs are compared by using ADR. In particular, it is established that the proposed SEs dominate the UMLE and that, near the null hypothesis, the RMLE is the most efficient. However, the RMLE is dominated by the UMLE as one moves far away from the restriction. These theoretical findings are confirmed by the simulation studies. Different real data sets are analyzed to illustrate the application of the proposed method.

5.2 Future Research

This dissertation has explored the generalization of some diffusion processes, but there are still many avenues of research that could be pursued in the future. This chapter details several potential areas of inquiry that could build upon the research presented in this dissertation. One area of future research could be to investigate the change point related problems, under the context of known and/or unknown numbers of change points. Chapter 2 outlined the parameter estimation and testing problem of GEOU process under the situation of one unknown change point. In the future, it will be important to consider cases where the number of change points is unknown. Under such a situation, the research could focus on the estimation of the number and location of the change points. This could be explored by combining loglikelihood method and least squared method and could yield valuable insights into improving the prediction accuracy and goodness-of-fit. The findings from GEOU process likewise suggest that the estimation of parameter dimensions is an important problem that deserves further investigations.

This could also be explored in the context of GCIR and GCKLS processes, and more accurate results could be expected. Further, it would be worthwhile to conduct research on detecting the existence of change points in GCIR and GCKLS processes. Another valuable area to explore is the implications of ergodicity for the GCKLS process with extended range of sensitivity parameter. This could involve the knowledge of transition probability and transition semi-group of diffusion processes. This could contribute to the proof of ergodicity of the solution to GCKLS model.

Furthermore, given the significant impact of simulation across various fields and its pivotal role in decision-making, experimentation, and problem-solving, a heightened emphasis will be placed on the utilization of simulation techniques in the chapter of simulation under different types of SDEs. Out-of-sample validation would be a good choice for testing how good the model is for predicting results on unseen new data. The dataset will be partitioned into two distinct segments: one designated as the training set for model development and the other as the testing set for evaluating model performance. The process will involve conducting k-fold cross-validation on the training data, a crucial step aimed at assessing the performance of each candidate model. This approach is instrumental in ensuring that the models generalize effectively when faced with unseen data. Evaluation methods such as SSE (Sum of Squared Error) will be employed to effectively compare the performance of candidate models using the validation data. It is important to note that model selection is an iterative procedure, necessitating a cautious approach to mitigate overfitting and underfitting while identifying the most appropriate model for the specific problem. Prediction with more steps ahead will be performed. More improvement could be expected under such a model selection procedure.

Appendices

A Proofs related to GEOU process

A.1 Proofs of parameter estimation and related problems

Proof of Proposition 2.3.18. The consistency of $\hat{\theta}_T(s)$ follows directly Proposition 2.3.6. Further, $\rho_T = \sigma T Q^{-1}(s,T)\frac{1}{\sqrt{T}}W(s,T)$ and then, by combining Proposition 2.3.17 and Slutsky's Theorem, $\rho_T \xrightarrow[T\to\infty]{D} \sigma\tilde{\Sigma}^{-1}\tilde{W}^* = \tilde{\rho} \sim \mathcal{N}_{2(p+1)}\left(0, \sigma^2\tilde{\Sigma}^{-1}\right)$. This completes the proof. □

Proof of Corollary 2.4.1. From (2.3.2) and (2.3.3), $U_{[0,T]}(p) = Q_{[0,T]}(p)\theta + \sigma W_{[0,T]}(p)$. By combining the triangle inequality and Cauchy-Schwartz inequalities,

$$\begin{aligned}
&||\log\mathcal{L}_N([0,T],\theta) - \log\mathcal{L}([0,T],\theta)||^{m/2} \\
&\leqslant 2^{m/2-1}\left(\frac{||\theta||}{\sigma}\right)^{m/2}\left\|\sum_{k=0}^{N-1}\Psi(t_k)(B_{t_{k+1}} - B_{t_k}) - \int_0^T \Psi(t)dB_t\right\|^{m/2} \\
&+2^{m/2-1}\left(\frac{||\theta||^2}{2\sigma^2}\right)^{m/2}\left\|\sum_{k=0}^{N-1}\Psi^\top(t_k)\Psi(t_k)\Delta_N - \int_0^T \Psi^\top(t)\Psi(t)dt\right\|^{m/2},
\end{aligned}$$

for any $\theta \in \Theta_0$. Then, since Θ_0 is a compact subset, Proposition 2.4.1 implies

$$||\log\mathcal{L}_N([0,T],\theta) - \log\mathcal{L}([0,T],\theta)||^{m/2} \leqslant 2^{m/2-1}M_0K(m,\Delta_N)O(T^{m/2}) + 2^{m/2-1}M_0\times$$

$$\sqrt{C_m \max\left\{(\Delta_N)^{m/2-1}3^{m-1}\left((\alpha\Delta_N)^{m/2} + o((\Delta_N)^{m/2})\right), (C_3(\Delta_N))^m(\Delta_N)^{m/2-1}\right\}T}$$

for some $M_0 > 0$. This completes the proof. □

Proof of Proposition 2.4.2. By (2.4.2),

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i)$$

$$= \frac{1}{T}\begin{bmatrix} \sum\limits_{i\in\mathbb{N}[0,T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i) & -\sum\limits_{i\in\mathbb{N}[0,T]} \varphi^\top(t_i)\ln X(t_i)(t_{i+1}-t_i) \\ -\sum\limits_{i\in\mathbb{N}[0,T]} \ln X_(t_i)\varphi(t_i)(t_{i+1}-t_i) & \sum\limits_{i\in\mathbb{N}[0,T]} (\ln X(t_i))^2(t_{i+1}-t_i) \end{bmatrix}.$$

Let $a = [a_1^\top, a_2]$ with a_1 a $p-$column vector, and a_2 a scalar. Then,

$$a\left[\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i)\right]a^\top = \sum_{i\in\mathbb{N}[0,T]} (a_1^\top\varphi(t_i) - a_2\ln X(t_i))^2(t_{i+1}-t_i) \geqslant 0,$$

and the equality holds if and only if $(a_1^\top\varphi(t_i) - a_2\ln X(t_i))^2 = 0$, almost everywhere for $i \in \mathbb{N}[0,T]$, which is $a_1^\top\varphi(t_i) - a_2\ln X(t_i) = 0$, almost everywhere for $i \in \mathbb{N}[0,T]$. This implies

$$\mathrm{P}\left(\omega : a_1^\top\varphi(t_i) - a_2(\ln X(t_i,\omega)) = 0, \forall\, i \in \mathbb{N}[0,T]\right) = 1. \tag{A.1}$$

Let us prove that if $a_1^\top\varphi(t_i) - a_2\ln X(t_i) = 0, \forall\, i \in \mathbb{N}[0,T]$ with probability 1, then, $a = \mathbf{0}_{(p+1)\times 1}$. Suppose that $a_2 \neq 0$, which means $a \neq \mathbf{0}_{(p+1)\times 1}$. From Proposition 2.2.1,

$$\ln X(t)|X_0 \sim \mathcal{N}(\mu(t,X_0), \sigma_0^2(t))$$

where $\mu(t,X_0) = \mathbb{E}[\ln X(t)|X_0]$, $\sigma_0^2(t) = \mathrm{Var}(\ln X(t)|X_0), \forall\, t \in [0,T]$. Then,

$$[a_1^\top\varphi(t_i) - a_2\ln X(t_i,\omega)]|X_0 \sim \mathcal{N}(a_1^\top\varphi(t_i) - a_2\mu(t_i,X_0), a_2^2\sigma_0^2(t_i)), \forall\, i \in \mathbb{N}[0,T].$$

Further, since $\alpha > 0$, from Proposition 2.2.1 and the independence between X_0 and Brownian motion $\{B_t, t \geqslant 0\}$ of Assumption 2.2,

$$\begin{aligned}\sigma_0^2(t_i) =& \mathrm{Var}\left(\ln X(t_i)|X_0\right) = \mathrm{Var}\left([e^{-\alpha t_i}(\ln X_0)\right.\\ &\left.+e^{-\alpha t_i}\sum_{k=1}^{p}\mu_k\int_0^{t_i} e^{\alpha s}\varphi_k(s)ds + \sigma e^{-\alpha t_i}\int_0^{t_i} e^{\alpha s}dB_s\right]\left|X_0\right)\end{aligned}$$

$$= \text{Var}\left(\sigma e^{-\alpha t_i}\int_0^{t_i} e^{\alpha s} dB_s | X_0\right) = \text{Var}\left(\sigma e^{-\alpha t_i}\int_0^{t_i} e^{\alpha s} dB_s\right)$$
$$= \sigma^2 e^{-2\alpha t_i}\text{Var}\left(\int_0^{t_i} e^{\alpha s} dB_s\right) = \sigma^2 e^{-2\alpha t_i}\left(\mathbb{E}\left[\left(\int_0^{t_i} e^{\alpha s} dB_s\right)^2\right] - \mathbb{E}^2\left[\int_0^{t_i} e^{\alpha s} dB_s\right]\right).$$

By Itô's isometry and the property of martingale,

$$\sigma_0^2(t_i) = \sigma^2 e^{-2\alpha t_i}\left\{\mathbb{E}\left[\int_0^{t_i} e^{2\alpha s} ds\right] - 0\right\} = \sigma^2 e^{-2\alpha t_i}\left(\frac{1}{2\alpha}(e^{2\alpha t_i} - 1)\right) = \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t_i}).$$

This implies that $\sigma_0^2(t_i) > 0$ for all $i \in \mathbb{N}[0, T]$. Thus, if $a_2 \neq 0$, $a_2^2\sigma_0^2(t_i) > 0$ for all $i \in \mathbb{N}[0, T]$. Then, $\text{P}\left(\left\{\omega : a_1^\top \varphi(t_i) - a_2(\ln X(t_i, \omega)) = 0, \forall\, i \in \mathbb{N}[0, T]\right\}\right) = 0$. This is a contradiction with (A.1). So, the assumption $a_2 \neq 0$ is not correct, which implies that $a_2 = 0$. From $a_1^\top \varphi(t_i) - a_2 \ln X(t_i, \omega) = 0$ in (A.1), $a_1^\top \varphi(t_i) = 0, \forall\, i \in \mathbb{N}[0, T]$. If $T \geqslant 1$, $[0, 1] \subset [0, T]$, by Assumption 2.2, $\{\varphi_1(t), \varphi_2(t), \dots, \varphi_p(t)\}$ is linearly independent on $[0, 1]$, this implies that $a_1^\top \varphi(t_i) = 0$ if and only if $a_1^\top = \mathbf{0}_{1\times(p+1)}$, $\forall\, i \in \mathbb{N}[0, T]$. Hence, if $T \geqslant 1$, the matrix $\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1} - t_i)$ is a positive definite matrix. Then, $\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1} - t_i)$ is positive definite. Second, by combining the triangle inequality,

$$\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1} - t_i} - \Sigma\right\|^{m/2} \leqslant 2^{m/2-1}\left(\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1} - t_i} - \frac{1}{T}Q_{[0,T]}\right\|^{m/2} + \left\|\frac{1}{T}Q_{[0,T]} - \Sigma\right\|^{m/2}\right),$$

By Proposition 2.3.4, and (2.4.5), $\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p)Z_i(p)/(t_{i+1} - t_i) - \Sigma\right\| \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{L^{m/2}} 0$, which completes the proof. □

Lemma A.1. *If at least one of the parameters, say,* μ_j, $(\mu_j \neq 0)$, $p_* < j \leqslant p_0$, *cannot be estimated, then for large* T, $\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\left(Z_i(p)(\theta(p) - \hat{\theta}(p_*+))\right)^2/(t_{i+1} - t_i) \geqslant \gamma_1|\mu_j|^2 > 0$*, with positive probability.*

Proof. If there is at least one parameter was not estimated, say μ_j, $(\mu_j \neq 0)$, $p_* < j \leqslant p$, since $\left(Z_i(p)(\theta(p)-\hat{\theta}(p_*+))\right)^2 \geqslant 0$ for each i, then,

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{\left(Z_i(p)(\theta(p)-\hat{\theta}(p_*+))\right)^2}{t_{i+1}-t_i} = (\theta(p)-\hat{\theta}(p_*+))^\top\left[\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1}-t_i}\right](\theta(p)-\hat{\theta}(p_*+))$$

$$\geqslant \gamma_1(T)\|\theta(p)-\hat{\theta}(p_*+)\|^2 = \gamma_1(T)\left(\sum_{j=1}^{p_*}(\hat{\mu}_j-\mu_j)^2+\sum_{j=p_*+1}^{p}(\mu_j-0)^2+(\hat{\alpha}-\alpha)^2\right).$$

This gives, $\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{\left(Z_i(p)(\theta(p)-\hat{\theta}(p_*+))\right)^2}{t_{i+1}-t_i} \geqslant \gamma_1(T)|\mu_j|^2$. By Corollary 2.4.2, $\gamma_1(T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \gamma_1$, with γ_1 strictly positive. Then, $\gamma_1(T)|\mu_j|^2 \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \gamma_1|\mu_j|^2 > 0$. This completes the proof. □

Proof of Proposition 2.4.4. By the triangle inequality,

$$\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,T]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}\right\| \leqslant \left\|\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,T]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}-\frac{1}{\sqrt{T}}\sigma W_{[0,T]}(p)\right\| + \frac{1}{\sqrt{T}}\left\|\sigma W_{[0,T]}(p)\right\|.$$

From (2.4.5), $\left\|\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,T]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}-\frac{1}{\sqrt{T}}\sigma W_{[0,T]}(p)\right\| \xrightarrow[\substack{T\to\infty\\ \Delta_N\to\infty}]{L^m} 0$, and then, from Proposition 2.4.3, assertion (2.4.8) holds. This completes the proof. □

Proof of Proposition 2.4.5. The proof will be completed by comparing the value of $\mathrm{IC}(p_0)$ and $\mathrm{IC}(p)$ for large T and small Δ_N. Following (2.4.6), for $p = p_0$,

$$\mathrm{IC}(p_0) = -2\left(\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i(p_0)\hat{\theta}(p_0)}{t_{i+1}-t_i}Y_i - \frac{1}{2\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{(Z_i(p_0)\hat{\theta}(p_0))^2}{t_{i+1}-t_i}\right) + (p_0+1)\log(\mathrm{N})$$

where $Z_i(p_0) = (\varphi_1(t_i), \varphi_2(t_i), \ldots, \varphi_{p_0}(t_i), -\ln X(t_i))(t_{i+1}-t_i)$, $\hat{\theta}(p_0) = (\hat{\mu}_1, \hat{\mu}_2, \ldots, \hat{\mu}_{p_0}, \hat{\alpha})^\top$.

$$\mathrm{IC}(p_0) = -2\left(\frac{1}{2\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left(-(Y_i-Z_i(p_0)\hat{\theta}(p_0))^2+(Y_i)^2\right)\right) + (p_0+1)\log(\mathrm{N}).$$

Since $Y_i = Z_i(p_0)\theta(p_0)+\epsilon_i$,

$$\mathrm{IC}(p_0) = -2\left(\frac{1}{2\sigma^2}\left(\sum_{i\in\mathbb{N}[0,T]}\frac{(Y_i)^2}{t_{i+1}-t_i} - \sum_{i\in\mathbb{N}[0,T]}\frac{\left(Z_i(p_0)\theta(p_0)+\epsilon_i-Z_i(p_0)\hat{\theta}(p_0)\right)^2}{t_{i+1}-t_i}\right)\right)$$

$$+ (p_0+1)\log(N).$$

Similarly, for $p \neq p_0$,

$$\mathrm{IC}(p) = -2\left(\frac{1}{2\sigma^2}\left(\sum_{i\in\mathbb{N}[0,T]}\frac{(Y_i)^2}{t_{i+1}-t_i} - \sum_{i\in\mathbb{N}[0,T]}\frac{\left(Z_i(p_0)\theta(p_0)+\epsilon_i - Z_i(p)\hat{\theta}(p)\right)^2}{t_{i+1}-t_i}\right)\right) + (p+1)\log(N).$$

For the case of $p = p^* > p_0$,

$$\begin{aligned}\mathrm{IC}(p^*) - \mathrm{IC}(p_0) = &-2\left\{\frac{1}{2\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left[\left(Z_i(p_0)\theta(p_0)+\epsilon_i - Z_i(p^*)\hat{\theta}(p^*)\right)^2\right.\right.\\ &\left.\left.-\left(Z_i(p_0)\theta(p_0)+\epsilon_i - Z_i(p_0)\hat{\theta}(p_0)\right)^2\right]\right\} + (p^*-p_0)\log(N).\end{aligned}$$

Note that, $\theta(p_0)$ and $\theta(p^*)$ have different dimensions. Let $\theta(p_0+) = (\mu_1, \mu_2, \ldots, \mu_{p_0}, 0_{p_0+1}, \ldots, 0_{p^*}, \alpha)^\top$ be a constructed auxiliary vector, with p^* as its dimension, $Z_i(p_0)\theta(p_0) - Z_i(p^*)\hat{\theta}(p^*) = Z_i(p^*)(\hat{\theta}(p_0+) - \theta(p^*))$. Then,

$$\begin{aligned}\mathrm{IC}(p^*) - \mathrm{IC}(p_0) = &-2\left\{\frac{1}{2\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left[\left(Z_i(p^*)(\hat{\theta}(p^*)-\theta(p_0+))\right)^2 + \epsilon_i^2\right.\right.\\ &-2\epsilon_i Z_i(p^*)\left(\hat{\theta}(p^*)-\theta(p_0+)\right) - \left(Z_i(p_0)(\hat{\theta}(p_0)-\theta(p_0))\right)^2 - \epsilon_i^2\\ &\left.\left.+2\epsilon_i Z_i(p_0)\left(\hat{\theta}(p_0)-\theta(p_0)\right)\right]\right\} + (p^*-p_0)\log(N)\\ = &-\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left[\left(Z_i(p^*)(\hat{\theta}(p^*)-\theta(p_0+))\right)^2 - \left(Z_i(p_0)(\hat{\theta}(p_0)-\theta(p_0))\right)^2\right.\\ &\left.+2\epsilon_i\left(Z_i(p_0)\hat{\theta}(p_0) - Z_i(p^*)\hat{\theta}(p^*)\right)\right] + (p^*-p_0)\log(N).\end{aligned}$$

By the previous estimation results, $\hat{\theta}(p^*) = \theta(p_0+) + \sigma Q^{-1}_{[0,T]}(p^*)W_{[0,T]}(p^*)$. Proposition 2.3.1 shows that $Q_{[0,T]}(p)$ is invertible provided that $T \geqslant 1$, which implies that $Q_{[0,T]}(p^*)$ is also invertible provided that $T \geqslant 1$. So by substituting $\hat{\theta}(p^*) = \theta(p_0+) + \sigma Q^{-1}_{[0,T]}(p^*)W_{[0,T]}(p^*)$ into $\mathrm{IC}(p^*) - \mathrm{IC}(p_0)$,

$$\begin{aligned}\mathrm{IC}(p^*) - \mathrm{IC}(p_0) = &-\frac{1}{\sigma^2}\left[\sigma^2 W^\top_{[0,T]}(p^*)Q^{-1}_{[0,T]}(p^*)\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p^*)Z_i(p^*)}{t_{i+1}-t_i}Q^{-1}_{[0,T]}(p^*)W_{[0,T]}(p^*)\right.\\ &-\sigma^2 W^\top_{[0,T]}(p_0)Q^{-1}_{[0,T]}(p_0)\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p_0)Z_i(p_0)}{t_{i+1}-t_i}Q^{-1}_{[0,T]}(p_0)W_{[0,T]}(p_0)\end{aligned}$$

$$+2\sigma \sum_{i\in\mathbb{N}[0,T]} \epsilon_i \Bigg(\frac{Z_i(p_0)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p_0) W_{[0,T]}(p_0)$$

$$- \frac{Z_i(p^*)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p^*) W_{[0,T]}(p^*) \Bigg) \Bigg] + (p^* - p_0)\log(N).$$

Further, from Proposition 2.3.4 and Proposition 2.4.2,

$$\frac{1}{T}\left\|Q_{[0,T]}(p_0)\right\| = O_p(1), \quad T\|Q^{-1}_{[0,T]}(p)\| = O_p(1), \quad \frac{1}{T}\left\|\sum_{i\in\mathbb{N}[0,T]} \frac{Z_i^\top(p) Z_i(p)}{\Delta_N}\right\| = O_p(1). \tag{A.2}$$

Hence, by Cauchy-Schwarz inequality,

$$\left\|W^\top_{[0,T]}(p^*) Q^{-1}_{[0,T]}(p^*) \sum_{i\in\mathbb{N}[0,T]} \frac{Z_i^\top(p^*) Z_i(p^*)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p^*) W_{[0,T]}(p^*)\right\|$$

$$\leqslant \left\|\frac{1}{\sqrt{T}} W_{[0,T]}(p^*)\right\|^2 \left\|T Q^{-1}_{[0,T]}(p^*)\right\|^2 \left\|\frac{1}{T} \sum_{i\in\mathbb{N}[0,T]} \frac{Z_i^\top(p^*) Z_i(p^*)}{t_{i+1}-t_i}\right\| = O_p(\log^{2a^*}(T)).$$

for some $0 < a^* < \boldsymbol{a}/2$. Similarly,

$$\left\|W^\top_{[0,T]}(p_0) Q^{-1}_{[0,T]}(p_0) \sum_{i\in\mathbb{N}[0,T]} \frac{Z_i^\top(p_0) Z_i(p_0)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p_0) W_{[0,T]}(p_0)\right\| = O_p(\log^{2a^*}(T)).$$

Further, from (2.4.7) for $p = p^*$, and Proposition 2.4.4 along with the triangle inequality, $\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,T]} \frac{\varepsilon_i Z_i(p^*)}{t_{i+1}-t_i}\right\| = O_p(\log^{a^*}(T))$. So,

$$\sigma \left|\sum_{i\in\mathbb{N}[0,T]} \left[\frac{\epsilon_i Z_i(p_0)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p_0) W_{[0,T]}(p_0) - \frac{\epsilon_i Z_i(p^*)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p^*) W_{[0,T]}(p^*)\right]\right|$$

$$\leqslant \sigma \left|\sum_{i\in\mathbb{N}[0,T]} \frac{\epsilon_i Z_i(p_0)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p_0) W_{[0,T]}(p_0)\right| + \sigma \left|\sum_{i\in\mathbb{N}[0,T]} \frac{\epsilon_i Z_i(p^*)}{t_{i+1}-t_i} Q^{-1}_{[0,T]}(p^*) W_{[0,T]}(p^*)\right|$$

$$\leqslant \frac{1}{\sqrt{T}} \left\|\sum_{i\in\mathbb{N}[0,T]} \frac{\epsilon_i Z_i(p_0)}{\Delta_N}\right\| T\|Q^{-1}_{[0,T]}(p_0)\| \frac{\sigma}{\sqrt{T}} \|W_{[0,T]}(p_0)\|$$

$$+ \frac{1}{\sqrt{T}} \left\|\sum_{i\in\mathbb{N}[0,T]} \frac{\epsilon_i Z_i(p^*)}{\Delta_N}\right\| T\|Q^{-1}_{[0,T]}(p^*)\| \frac{\sigma}{\sqrt{T}} \|W_{[0,T]}(p^*)\| = O_p(\log^{2a^*} T).$$

Therefore, for large T, $\mathrm{IC}(p^*) - \mathrm{IC}(p_0)$ is dominated by $(p^* - p_0)\log(N)$, which is positive. This implies that $\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(p^*) - \mathrm{IC}(p_0) > 0) = 1$ for all $p = p^* > p_0$.

For the case $p = p_* < p_0$,

$$\frac{1}{T}\left(\mathrm{IC}(p_*) - \mathrm{IC}(p_0)\right) = \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{\left(Y_i - Z_i(p_*)\hat{\theta}(p_*)\right)^2 - (Y_i)^2}{t_{i+1}-t_i} + \frac{(p_*+1)\log(N)}{T}$$

$$-\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{\left[\left(Y_i - Z_i(p_0)\hat{\theta}(p_0)\right)^2 - (Y_i)^2\right]}{t_{i+1}-t_i} + \frac{(p_0+1)\log(N)}{T}$$

$$= \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left[\left(Y_i - Z_i(p_*)\hat{\theta}(p_*)\right)^2 - \left(Y_i - Z_i(p_0)\hat{\theta}(p_0)\right)^2\right] + \frac{(p_*-p_0)\log(N)}{T}$$

$$= \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left[\left(Z_i(p_0)\theta(p_0) - Z_i(p_*)\hat{\theta}(p_*) + \epsilon_i\right)^2\right.$$

$$\left.-\left(Z_i(p_0)\theta(p_0) - Z_i(p_0)\hat{\theta}(p_0) + \epsilon_i\right)^2\right] + \frac{(p_*-p_0)\log(N)}{T}.$$

Further, $\frac{1}{T}\left(\mathrm{IC}(p_*) - \mathrm{IC}(p_0)\right)$ can be rewritten as

$$\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{1}{t_{i+1}-t_i}\left[\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+))\right)^2 - \left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)^2\right.$$

$$\left.+2\epsilon_i\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+)) - Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)\right] + \frac{(p_*-p_0)\log(N)}{T}.$$

Then, to simplify the notations, let

$$\frac{1}{T}\left(\mathrm{IC}(p_*) - \mathrm{IC}(p_0)\right) = (a_1(T) - a_2(T) + a_3(T)) + (p_* - p_0)\log(N)/T,$$

where

$$a_1(T) = \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+))\right)^2/(t_{i+1}-t_i),$$

$$a_2(T) = \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)^2/(t_{i+1}-t_i),$$

$$a_3(T) = \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}2\epsilon_i\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+)) - Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)/(t_{i+1}-t_i),$$

and $\hat{\theta}(p_*+)$ is defined in Lemma A.1. Since $p = p_* < p_0$, there is at least one parameter which can not be estimated. Without loss of generality, μ_k, $(\mu_k \neq 0)$, $p_* < k \leqslant p_0$ is supposed to be the parameter that cannot be consistently estimated. From Lemma A.1,

$$a_1(T) \geqslant \gamma_1 |\mu_k|^2/\sigma^2 > 0. \tag{A.3}$$

Further,

$$a_2(T) = \frac{1}{\sigma^2}(\theta(p_0) - \hat{\theta}(p_0))^\top \left[\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} Z_i^\top(p_0)Z_i(p_0)/(t_{i+1}-t_i)\right](\theta(p_0) - \hat{\theta}(p_0)).$$

From Proposition 2.3.6, and Proposition 2.4.2,

$$a_2(T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \frac{1}{\sigma^2}\mathbf{0}_{1\times(p_0+1)}\Sigma_{(p_0+1)\times(p_0+1)}\mathbf{0}_{(p_0+1)\times 1} = 0. \tag{A.4}$$

Let $\vec{a} = (0_1, \ldots, 0_{p_*}, \mu_{p_*+1}, \ldots, \mu_{p_0}, 0)^\top$. By using Proposition 2.4.4 and the fact that,

$$\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,T]}\frac{2\epsilon_i Z_i(p_0)}{t_{i+1}-t_i}\right\| = O_p(\log^{a*} T),$$

which implies that

$$\frac{1}{T}\left\|\sum_{i\in\mathbb{N}[0,T]}\frac{2\epsilon_i Z_i(p_0)}{t_{i+1}-t_i}\right\| \xrightarrow[T\to\infty]{\mathrm{P}} 0.$$

Meanwhile, as $\hat{\theta}(p_0) - \theta(p_0) \xrightarrow[T\to\infty]{a.s.} \mathbf{0}_{(p_0+1)\times 1}$,

$$\theta(p_0) - \hat{\theta}(p_*+) - (\theta(p_0) - \hat{\theta}(p_0)) \xrightarrow[T\to\infty]{a.s.} \vec{a},$$

with $\vec{a} = (0_1, \ldots, 0_{p_*}, \mu_{p_*+1}, \ldots, \mu_{p_0}, 0)^\top$, non-random,

$$\begin{aligned} a_3(T) = \;& \frac{1}{\sigma^2}\left[\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]} 2\epsilon_i Z_i(p_0)/(t_{i+1}-t_i)\right](\theta(p_0) - \hat{\theta}(p_*+) - (\theta(p_0) - \hat{\theta}(p_0))) \\ & \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \frac{1}{\sigma^2}\mathbf{0}\vec{a} = 0. \end{aligned} \tag{A.5}$$

For the last term, by Assumption 2.4,

$$(p_* - p_0)\log(N)/T \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0. \tag{A.6}$$

Therefore, by (A.3)-(A.6), $\lim\limits_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(p_*) - \mathrm{IC}(p_0) > 0) = 1$ for $p = p_* < p_0$. So, it concludes that for $p \neq p_0$, $\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(p) - \mathrm{IC}(p_0) > 0) = 1$.

Further, by Proposition 2.4.5,

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(p_0) \leq \mathrm{IC}(\hat{p})) = 1$$

and by definition of $\hat{p}$

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(p_0) \geq \mathrm{IC}(\hat{p})) = 1$$

This implies, for any $\epsilon > 0$,

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}(|\mathrm{IC}(p_0) - \mathrm{IC}(\hat{p})| > \epsilon) = 0$$

i.e. $\mathrm{IC}(\hat{p}) - \mathrm{IC}(p_0) \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$. Set $\hat{p} = \hat{p}_{(T)}$ and suppose $\hat{p}_{(T)} - p_0 \nrightarrow 0$ in probability, as $T \to \infty$. i.e. $\exists\, \epsilon_0 > 0, \forall\, T > 0, \exists\, T_0 > T$, such that $|\hat{p}_{(T_0)} - p_0| > \epsilon_0$, which implies that $\hat{p}_{(T_0)} > p_0 + \epsilon_0$ or $\hat{p}_{(T_0)} < p_0 - \epsilon_0$ hold. For $\hat{p}_{(T_0)} > p_0 + \epsilon_0$, there exists some $\eta_0 > 0$, such that $\hat{p}_{(T_0)} = p_0 + \epsilon_0 + \eta_0 > p_0 + \epsilon_0$, so $\mathrm{IC}(\hat{p}_{(T_0)}) = \mathrm{IC}(p_0 + \epsilon_0 + \eta_0)$. This is a contradiction with

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}\left(\mathrm{IC}(p_0) < \mathrm{IC}(p_0 + \epsilon_0 + \eta_0)\right) = 1$$

in Proposition 2.4.5. For $\hat{p}_{(T_0)} < p_0 - \epsilon_0$, there exists some $\eta_0 > 0$, such that $\hat{p}_{(T_0)} = p_0 - \epsilon_0 - \eta_0 < p_0 - \epsilon_0$, so $\mathrm{IC}(\hat{p}_{(T_0)}) = \mathrm{IC}(p_0 - \epsilon_0 - \eta_0)$. This is a contradiction with

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \mathrm{P}\left(\mathrm{IC}(p_0) < \mathrm{IC}(p_0 - \epsilon_0 - \eta_0)\right) = 1$$

in Proposition 2.4.5. So, $\hat{p} - p_0 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$. □

The following propositions and corollaries are important in proving that the proposed method improves the goodness-of-fit.

Proposition A.1. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then,*

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\varphi_{\hat{p}}(t_i) - \varphi_{p_0}(t_i)\right)(t_{i+1} - t_i)\mathbb{I}_{\{\hat{p}=p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0,$$

$$\left\| \frac{1}{T} \sum_{i=1}^{T} \left(\varphi_{\hat{p}}(t_i) - \varphi_{p_0+}(t_i) \right) (t_{i+1} - t_i) \mathbb{I}_{\{\hat{p} > p_0\}} \right\|^2, \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\varphi_{\hat{p}+}(t_i) - \varphi_{p_0}(t_i) \right) (t_{i+1} - t_i) \mathbb{I}_{\{\hat{p} < p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0.$$

Proof. The proof of the first claim is obvious. For the second claim, let $\delta > 0$.

$$\frac{1}{T} \sum_{i=1}^{N} \varphi_{\hat{p}}(t_i)(t_{i+1} - t_i) = \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_1(t_i)(t_{i+1} - t_i), \frac{1}{T} \sum_{i=1}^{N} \varphi_2(t_i)(t_{i+1} - t_i), \cdots, \frac{1}{T} \sum_{i=1}^{N} \varphi_{\hat{p}}(t_i) \right).$$

In the case $\hat{p} < p_0$, for $\forall\ \delta > 0$, Then,

$$\mathrm{P}\left(\left\| \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_{\hat{p}+}(t_i)(t_{i+1} - t_i) - \frac{1}{T} \sum_{i=1}^{N} \varphi_{p_0}(t_i)(t_{i+1} - t_i) \right) \mathbb{I}_{\{\hat{p} < p_0\}} \right\|^2 > \delta \right)$$

$$= \mathrm{P}\left(\left[\sum_{\hat{p}+1 \leqslant k \leqslant p_0} \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_k(t_i)(t_{i+1} - t_i) \right)^2 \right] \mathbb{I}_{\{\hat{p} < p_0\}} > \delta \right)$$

$$\leqslant \mathrm{P}\left(K_\varphi^2 \left[\sum_{\hat{p}+1 \leqslant k \leqslant p_0} 1 \right] \mathbb{I}_{\{\hat{p} < p_0\}} > \delta \right) \leqslant \mathrm{P}\left(K_\varphi^2 |\hat{p} - p_0| > \delta \right) = \mathrm{P}\left(|\hat{p} - p_0| > \frac{\delta}{K_\varphi^2} \right) \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{} 0.$$

Further,

$$\mathrm{P}\left(\left\| \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_{\hat{p}+}(t_i)(t_{i+1} - t_i) - \frac{1}{T} \sum_{i=1}^{N} \varphi_{p_0}(t_i)(t_{i+1} - t_i) \right) \mathbb{I}_{\{\hat{p} > p_0\}} \right\|^2 > \delta \right)$$

$$= \mathrm{P}\left(\left[\sum_{p_0+1 \leqslant k \leqslant \hat{p}} \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_k(t_i)(t_{i+1} - t_i) \right)^2 \right] \mathbb{I}_{\{\hat{p} > p_0\}} > \delta \right)$$

$$\leqslant \mathrm{P}\left(K_\varphi^2 \left[\sum_{p_0+1 \leqslant k \leqslant \hat{p}} 1 \right] \mathbb{I}_{\{\hat{p} > p_0\}} > \delta \right) \leqslant \mathrm{P}\left(K_\varphi^2 |\hat{p} - p_0| > \delta \right) = \mathrm{P}\left(|\hat{p} - p_0| > \frac{\delta}{K_\varphi^2} \right) \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{} 0.$$

This completes the proof. □

Corollary A.1. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then,*

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \left(Z_i(\hat{p}) - Z_i(p_0) \right) \mathbb{I}_{\{\hat{p} = p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

$$\left\| \frac{1}{T} \sum_{i=1}^{N} (Z_i(\hat{p}) - Z_i(p_0+)) \mathbb{I}_{\{\hat{p}>p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

$$\left\| \frac{1}{T} \sum_{i=1}^{N} (Z_i(\hat{p}+) - Z_i(p_0)) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0.$$

where $Z_i(\hat{p}+) = (\varphi_{\hat{p}+}(t_i), -\ln X(t_i))(t_{i+1} - t_i)$ *and* $Z_i(p_0+) = (\varphi_{p_0+}(t_i), -\ln X(t_i))(t_{i+1} - t_i)$.

Proof. The proof of the first claim is obvious. Further, by the definition of $Z_i(\hat{p}+)$ and Proposition A.1, for $\forall\, \delta > 0$, $\mathrm{P}\left(\left\| \frac{1}{T} \sum_{i=1}^{N} (Z_i(\hat{p}+) - Z_i(p_0)) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 > \delta \right)$

$$\begin{aligned} &\mathrm{P}\left(\left\| \frac{1}{T} \sum_{i=1}^{N} (Z_i(\hat{p}+) - Z_i(p_0)) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 > \delta \right) \\ &= \mathrm{P}\left(\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\varphi_{\hat{p}+}(t_i) - \varphi_{p_0}(t_i) \right) (t_{i+1} - t_i) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 > \delta \right) \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{} 0. \end{aligned} \tag{A.7}$$

Similarly, one proves that $\left\| \left(\frac{1}{T} \sum_{i=1}^{N} Z_i(\hat{p}+) - \frac{1}{T} \sum_{i=1}^{N} Z_i(p_0) \right) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0.$ □

Corollary A.2. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then,*

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\frac{Z_i^\top(\hat{p}) Z_i(\hat{p})}{t_{i+1} - t_i} - \frac{Z_i^\top(p_0) Z_i(p_0)}{t_{i+1} - t_i} \right) \mathbb{I}_{\{\hat{p}=p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\frac{Z_i^\top(\hat{p}) Z_i(\hat{p})}{t_{i+1} - t_i} - \frac{Z_i^\top(p_0+) Z_i(p_0+)}{t_{i+1} - t_i} \right) \mathbb{I}_{\{\hat{p}>p_0\}} \right\|^2, \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

$$\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\frac{Z_i^\top(\hat{p}+) Z_i(\hat{p}+)}{t_{i+1} - t_i} - \frac{Z_i^\top(p_0) Z_i(p_0)}{t_{i+1} - t_i} \right) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} \mathbf{0}.$$

Proof. For the first claim is obvious. Further, by the definition of $Z_i(\hat{p}+)$,

$$\begin{aligned} &\left\| \frac{1}{T} \sum_{i=1}^{N} \left(\frac{Z_i^\top(\hat{p}+) Z_i(\hat{p}+)}{t_{i+1} - t_i} - \frac{Z_i^\top(p_0) Z_i(p_0)}{t_{i+1} - t_i} \right) \mathbb{I}_{\{\hat{p}<p_0\}} \right\|^2 = \sum_{j=1}^{p_0} \sum_{k=\hat{p}+1}^{p_0} \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_j(t_i) \varphi_k(t_i)(t_{i+1} - t_i) \right)^2 \\ &\quad + \sum_{j=1}^{\hat{p}} \sum_{k=\hat{p}+1}^{p_0} \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_j(t_i) \varphi_k(t_i)(t_{i+1} - t_i) \right)^2 + 2 \sum_{k=\hat{p}+1}^{p_0} \left(\frac{1}{T} \sum_{i=1}^{N} \varphi_k(t_i)(\ln X(t_i))(t_{i+1} - t_i) \right)^2 \end{aligned}$$

$$\leqslant K_\varphi^4 p_0(p_0 - \hat{p}) + K_\varphi^4 \hat{p}(p_0 - \hat{p}) + 2K_\varphi^2(p_0 - \hat{p})\left(\frac{1}{T}\sum_{i=1}^{N}(\ln X(t_i))(t_{i+1} - t_i)\right)^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\text{P}} 0,$$

where the following convergence are used:

$$p_0 - \hat{p} \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\text{P}} 0 \;\; and \;\; \left(\frac{1}{T}\sum_{i=1}^{N}(\ln X(t_i))(t_{i+1} - t_i)\right)^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\text{P}} \left(\int_0^1 \tilde{r}(t)dt\right)^2.$$

Similarly, one can prove the third claim, this completes the proof. □

Proof of Proposition 2.4.6. By (2.4.6) and $T = N\Delta_N$,

$$\begin{aligned}
\frac{\text{IC}(p)}{T} =& \frac{-2}{T}\left(\frac{1}{2\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{-(Y_i - Z_i(p)\hat{\theta}(p))^2 + (Y_i)^2}{t_{i+1} - t_i}\right) + \frac{(p+1)\log(N)}{T} \\
=& \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{\left(Y_i - Z_i(p)\hat{\theta}(p)\right)^2}{\Delta_N} - \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{Y_i^2}{\Delta_N} + \frac{(p+1)\log(N)}{T}
\end{aligned}$$

and

$$\frac{1}{N}\text{SSE}(p) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{Y}_i(p)}{t_{i+1} - t_i} - \frac{Y_i}{t_{i+1} - t_i}\right)^2 = \frac{1}{T}\sum_{i=1}^{N}\frac{\left(Z_i(p)\hat{\theta}(p) - Y_i\right)^2}{\Delta_N},$$

which implies that

$$\frac{\text{IC}(p)}{T} = \frac{1}{\sigma^2}\frac{1}{N}\text{SSE}(p) - \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{Y_i^2}{\Delta_N} + \frac{(p+1)\log(N)}{T}. \tag{A.8}$$

By $\hat{p} = \arg\min_{p\in\mathbb{N}} \text{IC}(p)$ and by Proposition 2.4.5, together with (A.8) ,

$$\begin{aligned}
&\frac{1}{\sigma^2}\frac{1}{N}\text{SSE}(p) + \frac{(p+1)\log(N)}{T} - \left(\frac{1}{\sigma^2}\frac{1}{N}\text{SSE}(\hat{p}) + \frac{(\hat{p}+1)\log(N)}{T}\right) \\
&= \frac{1}{\sigma^2}\left(\frac{1}{N}\text{SSE}(p) - \text{SSE}(\hat{p})\right) + \frac{(p-\hat{p})\log(N)}{T} \geqslant 0,
\end{aligned}$$

with positive probability. From $\hat{p} - p \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\text{P}} 0$, and $\frac{\log(N)}{T} \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{} 0$,

$$\lim_{\substack{T\to\infty \\ \Delta_N\to 0}} \text{P}\left(\frac{1}{N}(\text{SSE}(p_0) - \text{SSE}(\hat{p})) \geqslant 0\right) = 1.$$

Next, it needs to show that, for $p \neq p_0$, $\frac{1}{N}(\text{SSE}(p) - \text{SSE}(p_0)) \geqslant 0$, with positive probability. First, if $p = p_* < p_0$,

$$\frac{1}{N}(\mathrm{SSE}(p) - \mathrm{SSE}(p_0)) = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{Y}_i(p_*)}{\Delta_N} - \frac{Y_i(p_0)}{\Delta_N}\right)^2 - \frac{1}{N}\sum_{i=1}^{N}\left(\frac{\hat{Y}_i(p_0)}{\Delta_N} - \frac{Y_i(p_0)}{\Delta_N}\right)^2.$$

Let $\hat{\theta}(p_*+)$ be as defined in Lemma A.1. Since $\frac{1}{N} = \frac{\Delta_N}{T}$,

$$\frac{1}{N}(\mathrm{SSE}(p) - \mathrm{SSE}(p_0)) = \frac{1}{T}\sum_{i=1}^{N}\left(\frac{\left(Z_i(p_*)\hat{\theta}(p_*) - Y_i(p_0)\right)^2}{\Delta_N} - \frac{\left(Z_i(p_0)\hat{\theta}(p_0) - Y_i(p_0)\right)^2}{\Delta_N}\right).$$

Similar to the proof of $p = p_* < p_0$ in Proposition 2.4.5,

$$\frac{1}{N}(\mathrm{SSE}(p_*) - \mathrm{SSE}(p_0)) = \frac{1}{T}\sum_{i=1}^{N}\frac{1}{\Delta_N}\left[\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+))\right)^2 - \left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)^2\right.$$
$$\left. + 2\epsilon_i\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+)) - Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)\right],$$

and then, to rewrite the notation,

$$\frac{1}{N}(\mathrm{SSE}(p_*) - \mathrm{SSE}(p_0)) = a_1(T) - a_2(T) + a_3(T),$$

where $\hat{\theta}(p_*+)$ is defined in Lemma A.1 and

$$\begin{aligned}
a_1(T) = &\ \tfrac{1}{T}\sum_{i=1}^{N}\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+))\right)^2/\Delta_N,\\
a_2(T) = &\ \tfrac{1}{T}\sum_{i=1}^{N}\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)^2/\Delta_N,\\
a_3(T) = &\ \tfrac{1}{T}\sum_{i=1}^{N}2\epsilon_i\left(Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_*+)) - Z_i(p_0)(\theta(p_0) - \hat{\theta}(p_0))\right)/\Delta_N.
\end{aligned}$$

Since $p = p_* < p_0$, there is at least one parameter which can not be estimated. Without loss of generality, μ_k, $(\mu_k \neq 0)$, $p_* < k \leqslant p_0$ is supposed the parameter, from Lemma A.1,

$$a_1(T) \geqslant \frac{\gamma_1|\mu_k|^2}{\sigma^2} > 0. \tag{A.9}$$

Further, by combining Proposition 2.3.6 and Proposition 2.4.2,

$$\begin{aligned}
a_2(T) &= \tfrac{1}{\sigma^2}(\theta(p_0) - \hat{\theta}(p_0))^\top\left[\tfrac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\tfrac{Z_i^\top(p_0)Z_i(p_0)}{\Delta_N}\right](\theta(p_0) - \hat{\theta}(p_0))\\
&\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \tfrac{1}{\sigma^2}\mathbf{0}'\Sigma\mathbf{0} = 0.
\end{aligned} \tag{A.10}$$

Further, let $\vec{a} = (0_1, \ldots, 0_{p_*}, \mu_{p_*+1}, \ldots, \theta_{p_0}, 0)^\top$. Using the fact that $\theta(p_0) - \hat{\theta}(p_*+) -$

$(\theta(p_0)-\hat{\theta}(p_0)) \xrightarrow[T\to\infty]{a.s.} \vec{a}_{(p_0+1)\times 1}$, and Proposition 2.4.4,

$$a_3(T)=\frac{1}{\sigma^2}\left[\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{2\epsilon_i Z_i(p_0)}{\Delta_N}\right](\theta(p_0)-\hat{\theta}(p_*+)-(\theta(p_0)-\hat{\theta}(p_0))) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \frac{1}{\sigma^2}\mathbf{0}'\vec{a}=0. \tag{A.11}$$

Therefore, by (A.9)-(A.11), $\lim\limits_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}\left(\frac{1}{N}(\mathrm{SSE}(p)-\mathrm{SSE}(p_0))\geqslant 0\right)=1$. Second, for the case of $p=p^*>p_0$.

$$\frac{1}{N}(\mathrm{SSE}(p^*)-\mathrm{SSE}(p_0))=\frac{1}{T}\sum_{i=1}^{N}\frac{\left(Z_i(p^*)\hat{\theta}(p^*)-Y_i(p_0)\right)^2}{\Delta_N}-\frac{1}{T}\sum_{i=1}^{N}\frac{\left(Z_i(p_0)\hat{\theta}(p_0)-Y_i(p_0)\right)^2}{\Delta_N}.$$

Since $Y_i(p_0)=Z_i(p_0)\theta(p_0)+\varepsilon_i$,

$$\begin{aligned}
\frac{1}{N}(\mathrm{SSE}(p^*)-\mathrm{SSE}(p_0))=&\frac{1}{T}\sum_{i=1}^{N}\left(Z_i(p^*)\left(\hat{\theta}(p^*)-\theta(p_0)\right)+\varepsilon_i\right)^2/\Delta_N\\
&-\frac{1}{T}\sum_{i=1}^{N}\left(Z_i(p_0)\left(\hat{\theta}(p_0)-\theta(p_0)\right)+\varepsilon_i\right)^2/\Delta_N\\
=&\left(\hat{\theta}(p^*)-\theta(p_0+)\right)^\top\frac{1}{T}\sum_{i=1}^{N}(Z_i^\top(p^*)Z_i(p^*)/\Delta_N)\left(\hat{\theta}(p^*)-\theta(p_0+)\right)\\
&-\frac{2}{T}\sum_{i=1}^{N}\frac{\varepsilon_i Z_i(p^*)}{\Delta_N}\left(\hat{\theta}(p^*)-\theta(p_0+)\right)\\
&-\left(\hat{\theta}(p_0)-\theta(p_0)\right)^\top\frac{1}{T}\sum_{i=1}^{N}(Z_i^\top(p_0)Z_i(p_0)/\Delta_N)\left(\hat{\theta}(p_0)-\theta(p_0)\right)\\
&+\frac{2}{T}\sum_{i=1}^{N}(\varepsilon_i Z_i(p_0)/\Delta_N)\left(\hat{\theta}(p_0)-\theta(p_0)\right).
\end{aligned}$$

Similar to the proof of $p=p^*>p_0$ in Proposition 2.4.5, one can prove that

$$\frac{2}{T}\sum_{i=1}^{N}(\varepsilon_i Z_i(p^*)/\Delta_N)\left(\hat{\theta}(p^*)-\theta(p_0+)\right)\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0,$$

$$\frac{2}{T}\sum_{i=1}^{N}(\varepsilon_i Z_i(p_0)/\Delta_N)\left(\hat{\theta}(p_0)-\theta(p_0)\right)\xrightarrow[T\to\infty]{\mathrm{P}} 0,$$

$$\left(\hat{\theta}(p_0)-\theta(p_0)\right)^\top\frac{1}{T}\sum_{i=1}^{N}(Z_i^\top(p_0)Z_i(p_0)/\Delta_N)\left(\hat{\theta}(p_0)-\theta(p_0)\right)\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0.$$

Then,

$$\left(\hat{\theta}(p^*) - \theta(p_0+)\right)^\top \frac{1}{T}\sum_{i=1}^{N} \frac{Z_i^\top(p^*)Z_i(p^*)}{\Delta_N}\left(\hat{\theta}(p^*) - \theta(p_0)\right)$$

$$\geqslant \gamma_{\min}\left(\hat{\theta}(p^*) - \theta(p_0+)\right)^\top\left(\hat{\theta}(p^*) - \theta(p_0+)\right) \geqslant 0,$$

where $\gamma_{\min}$ is the smallest eigenvalue of the matrix $\frac{1}{T}\sum_{i=1}^{N} Z_i^\top(p^*)Z_i(p^*)/\Delta_N$. Note that, for the case of $p^* > p_0$,

$$\left(\hat{\theta}(p^*) - \theta(p_0+)\right)^\top \frac{1}{T}\sum_{i=1}^{N}(Z_i^\top(p^*)Z_i(p^*)/\Delta_N)\left(\hat{\theta}(p^*) - \theta(p_0)\right) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0.$$

This implies that

$$\lim_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}\left(\frac{1}{N}(\mathrm{SSE}(p) - \mathrm{SSE}(p_0)) \geqslant 0\right) = 1.$$

This completes the proof. □

Proposition A.2. *Suppose that Assumption 2.1-2.3 and Assumption 2.4 hold,*

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(Z_i^\top(p_0)Z_i(p_0))/(t_{i+1} - t_i), \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} s\Sigma_1, \quad \text{(A.12)}$$

$$\frac{1}{T}\sum_{i\in\mathbb{N}[sT,T]} Z_i^\top(p_0)Z_i(p_0)/(t_{i+1} - t_i) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} (s\Sigma_1, (1-s)\Sigma_2)\,. \quad \text{(A.13)}$$

Proof. By triangle inequality,

$$\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1} - t_i} - s\Sigma_1\right\|^{m/2} \leqslant 2^{m/2}\left(\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1} - t_i} - \frac{1}{T}Q_{[0,sT]}\right\|^{m/2} + \left\|\frac{1}{T}Q_{[0,sT]} - s\Sigma_1\right\|^{m/2}\right).$$

By Proposition 2.3.17 along with the relation (2.4.5), by replacing the interval $[0, T]$ by $[0, sT]$,

$$\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}\frac{Z_i^\top(p)Z_i(p)}{t_{i+1} - t_i} - s\Sigma_1\right\| \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} 0,$$

which implies that (A.12) hold. Similarly, (A.13) can be proven. This completes the proof. □

Let $\gamma_1(s,T)$ be the smallest eigenvalue of matrix $\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i)$ and $\gamma_2(s,T)$ be the smallest eigenvalue of matrix $\frac{1}{T}\sum_{i\in\mathbb{N}[sT,T]} Z_i^\top(p)Z_i(p)/(t_{i+1}-t_i)$ and let γ_k be the smallest eigenvalue of Σ_k, for $k=1,2$.

Lemma A.2. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then* $\gamma_1(s,T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} s\gamma_1$ *and* $\gamma_2(s,T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} (1-s)\gamma_2$.

Proof. By combining Proposition A.3, Proposition A.2 and Corollary 2.4.2 along with the fact that Σ_1 is a positive definite matrix, $\gamma_1(s,T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \gamma_1(s) = s\gamma_1$ with γ_1 strictly positive. Similarly, $\gamma_2(s,T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} (1-s)\gamma_2$. □

Let $\hat{\theta}^{(k)}(p_*+) = (\hat{\mu}_1{}^{(k)}, \hat{\mu}_2^{(k)}, \cdots, \hat{\mu}_{p_*}^{(k)}, 0_{p_*+1}, \cdots, 0_p, \hat{\alpha}^{(k)})^\top$, $k=1,2$. The following lemma is useful in proving that the information criterion $\mathrm{IC}(c,p)$ reaches its minimum value at the exact dimension (c^0, p_0).

Lemma A.3. *Suppose that Assumption 2.1-2.3 and Assumption 2.4 hold and suppose that at least one of the parameters, say* $\mu_j^{(k)}$, $(\mu_j^{(k)} \neq 0)$, $p_* < j \leqslant p, k=1,2$, *cannot be consistently estimated, then for large* T,

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]} \frac{1}{t_{i+1}-t_i}\left(Z_i(p)(\theta^{(1)}(p)-\hat{\theta}^{(1)}(p_*+))\right)^2/(t_{i+1}-t_i) \geqslant s\gamma_1\left|\mu_j^{(1)}\right|^2 > 0,$$

$$\frac{1}{T}\sum_{i\in\mathbb{N}[sT,T]} \left(Z_i(p)(\theta^{(2)}(p)-\hat{\theta}^{(2)}(p_*+))\right)^2/(t_{i+1}-t_i) \geqslant (1-s)\gamma_2\left|\mu_j^{(2)}\right|^2 > 0,$$

with positive probability.

Proof. For the process on the observed interval $[0, sT]$, if there is at least one parameter was not consistently estimated, say $\mu_j^{(1)}$, $(\mu_j^{(1)}\neq 0)$, $p_* < j \leqslant p$.

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]} \left(Z_i(p)(\theta^{(1)}(p)-\hat{\theta}^{(1)}(p_*+))\right)^2/(t_{i+1}-t_i)$$

$$= (\theta^{(1)}(p) - \hat{\theta}^{(1)}(p_*+))^\top \left[\frac{1}{T} \sum_{i \in \mathbb{N}[0,sT]} Z_i^\top(p) Z_i(p)/(t_{i+1} - t_i) \right] (\theta^{(1)}(p) - \hat{\theta}^{(1)}(p_*+))$$

$$\geqslant \gamma_1(s,T) \|\theta^{(1)}(p) - \hat{\theta}^{(1)}(p_*+)\|^2$$

$$= \gamma_1(s,T) \left(\sum_{j=1}^{p_*} (\hat{\mu}_j^{(1)} - \mu_j^{(1)})^2 + \sum_{j=p_*+1}^{p_0} (\mu_j^{(1)} - 0)^2 + (\hat{\alpha}^{(1)} - \alpha^{(1)})^2 \right).$$

Then, $\frac{1}{T} \sum_{i \in \mathbb{N}[0,sT]} \left(Z_i(p)(\theta^{(1)}(p) - \hat{\theta}^{(1)}(p_*+)) \right)^2 \geqslant \gamma_1(s,T) \left| \mu_j^{(1)} \right|^2, \ p_* \leqslant j \leqslant p_0$. By the proof of Lemma A.2, $\gamma_1(s,T) \left| \mu_j^{(1)} \right|^2 \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\text{P}} s\gamma_1 \left| \mu_j^{(1)} \right|^2 > 0$. By using the same techniques, the second inequality is proven. This completes the proof. □

Before presenting this important result, The following propositions and lemmas are derived, which play a crucial role in establishing this result. Proposition A.3-Proposition A.5 and Lemma A.2, Lemma A.3 are based on the exact rate of the change-point s^0, while Proposition A.6-Proposition A.9 are based on $\hat{s}$ which is an estimator of s.

Proposition A.3. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then,* $\frac{1}{T} \sum_{i \in \mathbb{N}[a,b]} (Z_i^\top(p) Z_i(p)/(t_{i+1} - t_i)$ *is a positive definite matrix for* $b - a \geqslant 1$.

The proof follows from the proof of Proposition 2.4.2. Next, to emphasize the parameter dimension p in our notations, let $W(s,T,p)$ be the vector $W(s,T)$ with dimension p and $Q(s,T,p)$ be the matrix $Q(s,T,p)$ with size $2(p+1) \times 2(p+1)$. Let $W_{[0,sT]}(p)$ be the vector $W_{[0,sT]}(s)$ with dimension p and $Q_{[0,sT]}(p)$ be the matrix $Q_{[0,sT]}$ with size $(p+1) \times (p+1)$. The following proposition shows that, for some $0 < a^* < \boldsymbol{a}/2$, $\left(\frac{1}{\sqrt{T}} \|W(s,T,p)\| \right) / \left(\log^{a^*}(T) \right)$ is bounded in probability.

Proposition A.4. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then, for some* $0 < a^* < \boldsymbol{a}/2$,

$$\frac{1}{\sqrt{T}} \|W(s,T,p)\| = O_p(\log^{a^*}(T)).$$

Proof. The proof follows from the Proposition 2.4.3. □

Proposition A.5. *Suppose that Assumption 2.1-2.3 and Assumption 2.4 hold, then, for some* $0 < a^* < \boldsymbol{a}/2$*,* $\forall p \geqslant 1$*,*

$$\frac{1}{\sqrt{T}}\left(\left\|\sum_{i\in\mathbb{N}[0,sT]}\varepsilon_i Z_i(p)/(t_{i+1}-t_i)\right\|,\left\|\sum_{i\in\mathbb{N}[sT,T]}\varepsilon_i Z_i(p)/(t_{i+1}-t_i)\right\|\right) = O_p(\log^{a^*}(T)). \quad \text{(A.14)}$$

Proof. By combining the triangle inequality,

$$\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,sT]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}\right\| \leqslant \frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,sT]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}-\sigma W_T(s,p)\right\| + \frac{1}{\sqrt{T}}\left\|\sigma W_T(s,p)\right\|.$$

From the relation Equation (2.4.5) along with $\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,sT]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}-\sigma W_T(s,p)\right\| \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} 0$, which implies that $\frac{1}{\sqrt{T}}\left\|\sum_{i\in\mathbb{N}[0,sT]}\frac{\varepsilon_i Z_i(p)}{t_{i+1}-t_i}-\sigma W_T(s,p)\right\| \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0$. Further, from Proposition A.4, $\frac{\sigma}{\sqrt{T}}\|W_T(s,p)\| = O_p(\log^{a^*} T)$ is proven. This implies that the assertion in (A.14). Further, by using the same techniques, the second statement in (A.14) is proven. This completes the proof. □

Proof of Proposition 2.4.7. Let $\mathrm{SSE}_1 = \mathrm{SSE}(\hat{s},\hat{p},\hat{\boldsymbol{\theta}})/N$, and $\mathrm{SSE}_2 = \mathrm{SSE}(s^0,p_0,\hat{\boldsymbol{\theta}}_0)/N$, where $\hat{\boldsymbol{\theta}}$ is the estimator based on the estimation of the change point and $\hat{\boldsymbol{\theta}}_0$ is the estimator based on the true change point, denoted by s^0. Since $\mathrm{SSE}_1 \leqslant \mathrm{SSE}_2$ with probability 1, it remains to show that if the change point is not consistently estimated, $\mathrm{SSE}_1 > \mathrm{SSE}_2$ with positive probability yielding a contradiction. Indeed, if the change point is not consistently estimated, $|\hat{s}T - s^0T| > \eta T$, for some constant $0 < \eta < 1$. Without loss of generality, $0 < s^0T < \hat{s}T < T$ is supposed to be satisfied. Let $\hat{Y}_i(p,s)$ be the predicted value of Y_i based on the parameter p and s and $\hat{\boldsymbol{\theta}}(p,s)$ be the estimator of

$\boldsymbol{\theta}$ based on the parameter p and s. Then,

$$\mathrm{SSE}_1 - \mathrm{SSE}_2 = \left(\frac{1}{N}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\hat{Y}_i(\hat{p},\hat{s})}{\Delta_N} - \frac{Y_i}{\Delta_N}\right)^2 - \frac{1}{N}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\hat{Y}_i(p_0,s^0)}{\Delta_N} - \frac{Y_i}{\Delta_N}\right)^2\right)\mathbb{I}_{\{\hat{p}>p_0\}} \tag{A.15}$$

$$+\left(\frac{1}{N}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\hat{Y}_i(\hat{p},\hat{s})}{\Delta_N} - \frac{Y_i}{\Delta_N}\right)^2 - \frac{1}{N}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\hat{Y}_i(p_0,s^0)}{\Delta_N} - \frac{Y_i}{\Delta_N}\right)^2\right)\mathbb{I}_{\{\hat{p}=p_0\}} \tag{A.16}$$

$$+\left(\frac{1}{N}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\hat{Y}_i(\hat{p},\hat{s})}{\Delta_N} - \frac{Y_i}{\Delta_N}\right)^2 - \frac{1}{N}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\hat{Y}_i(p_0,s^0)}{\Delta_N} - \frac{Y_i}{\Delta_N}\right)^2\right)\mathbb{I}_{\{\hat{p}<p_0\}}. \tag{A.17}$$

Note that $T = N\Delta_N, Y_i = Z_i(p_0)\boldsymbol{\theta}(p_0) + \varepsilon_i$,

$$(A.15) = \frac{\Delta_N}{T}\sum_{i\in\mathbb{N}[0,T]}\left(\left(\frac{Z_i(\hat{p},\hat{s})}{\Delta_N}(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+))\right)^2 - \left(\frac{Z_i(p_0,s^0)}{\Delta_N}(\hat{\theta}(p_0,s^0) - \boldsymbol{\theta}(p_0))\right)^2\right)\mathbb{I}_{\{\hat{p}>p_0\}}$$

$$-\frac{2\Delta_N}{T}\sum_{i\in\mathbb{N}[0,T]}\left[\left(\varepsilon_i\frac{Z_i(\hat{p},\hat{s})}{\Delta_N}(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+))\right) + \left(\varepsilon_i\frac{Z_i(p_0,s^0)}{\Delta_N}(\hat{\theta}(p_0,s^0) - \boldsymbol{\theta}(p_0))\right)\right]\mathbb{I}_{\{\hat{p}>p_0\}}$$

$$= (\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+))^\top\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{\mathcal{Z}_i^\top(\hat{p})\mathcal{Z}_i(\hat{p})}{\Delta_N}(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+))\mathbb{I}_{\{\hat{p}>p_0\}}$$

$$-(\hat{\boldsymbol{\theta}}(p_0,s_0) - \boldsymbol{\theta}(p_0))^\top\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{\mathcal{Z}_i^\top(p_0)\mathcal{Z}_i(p_0)}{\Delta_N}(\hat{\boldsymbol{\theta}}(p_0,s_0) - \boldsymbol{\theta}(p_0))\mathbb{I}_{\{\hat{p}>p_0\}}$$

$$-\frac{2}{T}\sum_{i\in\mathbb{N}[0,T]}\left(\varepsilon_i\mathcal{Z}_i(\hat{p})(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+))\right)\mathbb{I}_{\{\hat{p}>p_0\}} + \frac{2}{T}\sum_{i\in\mathbb{N}[0,T]}\left(\varepsilon_i\mathcal{Z}_i(p_0)(\hat{\boldsymbol{\theta}}(p_0,s_0) - \boldsymbol{\theta}(p_0))\right)\mathbb{I}_{\{\hat{p}>p_0\}}.$$

Since s is not consistently estimated, and from $\left(\mathcal{Z}_i(\hat{p})(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+)\right)^2\mathbb{I}_{\{\hat{p}>p_0\}} \geqslant 0$,

$$\frac{1}{T}\left(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+)\right)^\top\sum_{i\in\mathbb{N}[0,T]}\frac{\mathcal{Z}_i^\top(\hat{p})\mathcal{Z}_i(\hat{p})}{\Delta_N}\left(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s}) - \boldsymbol{\theta}(p_0+)\right)\mathbb{I}_{\{\hat{p}>p_0\}}$$

$$\leqslant \eta\left((\boldsymbol{\mu}_1(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))^\top\frac{1}{\eta T}\sum_{i\in\mathbb{N}[s^0T-\eta T,s^0T]}\frac{Z_{1i}^\top(\hat{p})Z_{1i}(\hat{p})}{\Delta_N}(\boldsymbol{\mu}_1(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))\right)\mathbb{I}_{\{\hat{p}>p_0\}}$$
$$+\eta\left((\boldsymbol{\mu}_2(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p}))^\top\frac{1}{\eta T}\sum_{i\in\mathbb{N}[s^0T,s^0T+\eta T]}\frac{Z_{2i}^\top(\hat{p})Z_{2i}(\hat{p})}{\Delta_N}(\boldsymbol{\mu}_2(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))\right)\mathbb{I}_{\{\hat{p}>p_0\}}. \tag{A.18}$$

Let $\gamma_{11}(T), \gamma_{12}(T)$ be the smallest eigenvalues of the matrices $\displaystyle\sum_{i\in\mathbb{N}[s^0T-\eta T,s^0T]}\frac{Z_{1i}^\top(\hat{p})Z_{1i}(\hat{p})}{\Delta_N\eta T}$

and $\sum_{i\in\mathbb{N}[s^0T,s^0T+\eta T]} \dfrac{Z_{2i}^\top(\hat{p})Z_{2i}(\hat{p})}{\Delta_N \eta T}$, respectively, while $\hat{p} > p_0$. Then,

$$(A.18) \geqslant \eta \min(\gamma_{11}(T), \gamma_{12}(T)) \left(\|\boldsymbol{\mu}_1(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p}, \hat{s})\|^2 + \|\boldsymbol{\mu}_2(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p}, \hat{s})\|^2 \right) \mathbb{I}_{\{\hat{p}>p_0\}}.$$

Using the convexity of a quadratic function,

$\|\boldsymbol{\mu}_1(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p}, \hat{s})\|^2 + \|\boldsymbol{\mu}_2(p_0+) - \hat{\boldsymbol{\mu}}_1(\hat{p}, \hat{s})\|^2 \geqslant \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 /2$. Hence,

$$\frac{\Delta_N}{T} \sum_{i\in\mathbb{N}[0,T]} \left(\frac{\mathcal{Z}_i(\hat{p})}{\Delta_N} (\hat{\boldsymbol{\theta}}(\hat{p}, \hat{s}) - \boldsymbol{\theta}(p_0+)) \right)^2 \geqslant \eta \frac{\min(\gamma_{11}(T), \gamma_{12}(T))}{2} \|\boldsymbol{\theta}_1(p_0) - \boldsymbol{\theta}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}>p_0\}}.$$

By Proposition A.10, $\gamma_{11}(T)$ and $\gamma_{12}(T)$ are both bounded away from 0 and $\eta \min\{\gamma_{11}(T), \gamma_{12}(T)\}$ is also bounded away from 0. Therefore, the right-hand side of the inequality $\eta \min\{\gamma_{11}(T), \gamma_{12}(T)\} /2 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2$ is strictly positive. Further, since $\hat{\boldsymbol{\theta}}(p_0, s_0) - \boldsymbol{\theta}(p_0) \xrightarrow[T\to\infty]{a.s.} 0$,

$$-\left(\hat{\boldsymbol{\theta}}(p_0, s_0) - \boldsymbol{\theta}(p_0)\right)^\top \sum_{i\in\mathbb{N}[0,T]} \frac{\mathcal{Z}_i^\top(p_0)\mathcal{Z}_i(p_0)}{\Delta_N} \left(\hat{\boldsymbol{\theta}}(p_0, s_0) - \boldsymbol{\theta}(p_0)\right) \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0.$$

Proposition A.4 and Proposition A.9 imply that $\frac{2}{T} \left\| \sum_{i\in\mathbb{N}[0,T]} \varepsilon_i \mathcal{Z}_i(\hat{p}) \right\| \mathbb{I}_{\{\hat{p}>p_0\}} = o_p(1)$ and $\frac{2}{T} \left\| \sum_{i\in\mathbb{N}[0,T]} \varepsilon_i \mathcal{Z}_i(p_0) \right\| \mathbb{I}_{\{\hat{p}>p_0\}} = o_p(1)$. So,

$$-\frac{2}{T} \sum_{i\in\mathbb{N}[0,T]} \left(\varepsilon_i \mathcal{Z}_i(\hat{p}) \left(\hat{\boldsymbol{\theta}}(\hat{p}, \hat{s}) - \boldsymbol{\theta}(p_0+) \right) \right) \mathbb{I}_{\{\hat{p}>p_0\}}$$
$$+\frac{2}{T} \sum_{i\in\mathbb{N}[0,T]} \left(\varepsilon_i \mathcal{Z}_i(p_0) \left(\hat{\boldsymbol{\theta}}(p_0, s_0) - \boldsymbol{\theta}(p_0) \right) \right) \mathbb{I}_{\{\hat{p}>p_0\}} \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0.$$

This is for large T,

$$(A.15) \geqslant C_1 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}>p_0\}} \tag{A.19}$$

with positive probability, where $C_1 = \lim_{T\to\infty} \min(s\gamma_{11}(T), (1-s)\gamma_{12}(T))/2 > 0$. Further,

$$(A.16) = \left(\hat{\boldsymbol{\theta}}(\hat{p}, \hat{s}) - \boldsymbol{\theta}(p_0+)\right)^\top \frac{1}{T} \sum_{i\in\mathbb{N}[0,T]} \frac{\mathcal{Z}_i^\top(\hat{p})\mathcal{Z}_i(\hat{p})}{\Delta_N} \left(\hat{\boldsymbol{\theta}}(\hat{p}, \hat{s}) - \boldsymbol{\theta}(p_0+)\right) \mathbb{I}_{\{\hat{p}=p_0\}}$$

$$-\left(\hat{\boldsymbol{\theta}}(p_0, s_0)-\boldsymbol{\theta}(p_0)\right)^\top \frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{\mathcal{Z}_i^\top(p_0)\mathcal{Z}_i(p_0)}{\Delta_N}\left(\hat{\boldsymbol{\theta}}(p_0, s_0)-\boldsymbol{\theta}(p_0)\right)\mathbb{I}_{\{\hat{p}=p_0\}}$$

$$-\frac{2}{T}\sum_{i\in\mathbb{N}[0,T]}\left(\varepsilon_i\mathcal{Z}_i(\hat{p})\left(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s})-\boldsymbol{\theta}(p_0+)\right)\right)\mathbb{I}_{\{\hat{p}=p_0\}}$$

$$+\frac{2}{T}\sum_{i\in\mathbb{N}[0,T]}\left(\varepsilon_i\mathcal{Z}_i(p_0)\left(\hat{\boldsymbol{\theta}}(p_0,s_0)-\boldsymbol{\theta}(p_0)\right)\right)\mathbb{I}_{\{\hat{p}=p_0\}}.$$

Since s is not consistently estimated, and from $\left(\mathcal{Z}_i(\hat{p})(\hat{\theta}(\hat{p},\hat{s})-\boldsymbol{\theta}(p_0))\right)^2\mathbb{I}_{\{\hat{p}=p_0\}}\geqslant 0$,

$$\left(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s})-\boldsymbol{\theta}(p_0)\right)^\top\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{\mathcal{Z}_i^\top(\hat{p})\mathcal{Z}_i(\hat{p})}{\Delta_N}\left(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s})-\boldsymbol{\theta}(p_0+)\right)\mathbb{I}_{\{\hat{p}=p_0\}}$$

$$\begin{aligned}\geqslant \eta&\left((\boldsymbol{\mu}_1(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))^\top\frac{1}{\eta T}\sum_{i\in\mathbb{N}[s^0T-\eta T,s^0T]}\frac{Z_{1i}^\top(\hat{p})Z_{1i}(\hat{p})}{\Delta_N}(\boldsymbol{\mu}_1(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))\right)\mathbb{I}_{\{\hat{p}=p_0\}}\\ +\eta&\left((\boldsymbol{\mu}_2(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))^\top\frac{1}{\eta T}\sum_{i\in\mathbb{N}[s^0T,s^0T+\eta T]}\frac{Z_{2i}^\top(\hat{p})Z_{2i}(\hat{p})}{\Delta_N}(\boldsymbol{\mu}_2(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s}))\right)\mathbb{I}_{\{\hat{p}=p_0\}}.\end{aligned} \tag{A.20}$$

Let $\gamma_{21}(T),\gamma_{22}(T)$ be the smallest eigenvalues of the matrices $\sum_{i\in\mathbb{N}[s^0T-\eta T,s^0T]}\frac{Z_{1i}^\top(p_0)Z_{1i}(p_0)}{\Delta_N\eta T}$ and $\sum_{i\in\mathbb{N}[s^0T,s^0T+\eta T]}\frac{Z_{2i}^\top(p_0)Z_{2i}(p_0)}{\Delta_N\eta T}$, respectively. Then,

$$\begin{aligned}(A.20)&\geqslant\eta\gamma_{21}(T)\left\|\boldsymbol{\mu}_1(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s})\right\|^2+\eta\gamma_{22}(T)\left\|\boldsymbol{\mu}_2(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s})\right\|^2\mathbb{I}_{\{\hat{p}=p_0\}}\\&\geqslant\eta\min(\gamma_{21}(T),\gamma_{22}(T))\left(\left\|\boldsymbol{\mu}_1(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s})\right\|^2+\left\|\boldsymbol{\mu}_2(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s})\right\|^2\right)\mathbb{I}_{\{\hat{p}=p_0\}}.\end{aligned}$$

Using the convexity of a quadratic function,

$$\left\|\boldsymbol{\mu}_1(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s})\right\|^2+\left\|\boldsymbol{\mu}_2(p_0)-\hat{\boldsymbol{\mu}}_1(\hat{p},\hat{s})\right\|^2\geqslant\left\|\boldsymbol{\mu}_1(p_0)-\boldsymbol{\mu}_2(p_0)\right\|^2/2.$$

Hence,

$$\frac{\Delta_N}{T}\sum_{i\in\mathbb{N}[0,T]}\left(\frac{\mathcal{Z}_i(\hat{p})}{\Delta_N}(\hat{\boldsymbol{\theta}}(\hat{p},\hat{s})-\boldsymbol{\theta}(p_0))\right)^2\mathbb{I}_{\{\hat{p}=p_0\}}$$

$$\geqslant\frac{\eta\min(\gamma_{21}(T),\gamma_{22}(T))}{2}\left\|\boldsymbol{\mu}_1(p_0)-\boldsymbol{\mu}_2(p_0)\right\|^2\mathbb{I}_{\{\hat{p}=p_0\}}.$$

By Proposition A.10, $\gamma_{21}(T)$ and $\gamma_{22}(T)$ are both bounded away from 0. This implies that $\eta\min(\gamma_{21}(T),\gamma_{22}(T))$ is also bounded away from 0. Therefore, the right-hand side of the inequality $\eta \min\left(\gamma_{21}(T), \gamma_{22}(T)\right)/2\, \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2$ is positive. Similarly,

$$(\hat{\theta}(\hat{p}, s^0) - \boldsymbol{\theta}(p_0))^\top \frac{1}{T} \sum_{i \in \mathbb{N}[0,T]} \frac{Z_i^\top(\hat{p}) Z_i(\hat{p})}{\Delta_N} (\hat{\theta}(\hat{p}, s^0) - \boldsymbol{\theta}(p_0)) \mathbb{I}_{\{\hat{p}=p_0\}} \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

$$-\frac{2}{T} \sum_{i \in \mathbb{N}[0,T]} \left[\left(\varepsilon_i \mathcal{Z}_i(\hat{p})(\hat{\boldsymbol{\theta}}(\hat{p}, \hat{s}) - \boldsymbol{\theta}(p_0)) \right) \mathbb{I}_{\{\hat{p}=p_0\}} + \left(\varepsilon_i \mathcal{Z}_i(\hat{p})(\hat{\theta}(\hat{p}, s^0) - \boldsymbol{\theta}(p_0)) \right) \mathbb{I}_{\{\hat{p}=p_0\}} \right] \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{\mathrm{P}} 0,$$

which implies that for large T,

$$(A.16) \geqslant C_2 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}=p_0\}}. \tag{A.21}$$

with a positive probability, where $C_2 = \lim\limits_{\substack{T\to\infty \\ \Delta_N \to 0}} \eta \min(\gamma_{12}(T), \gamma_{22}(T))/2 > 0$. Similarly to the proof of (A.19) and (A.21), for large T,

$$(A.17) \geqslant C_3 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}<p_0\}} \tag{A.22}$$

with positive probability, where $C_3 = \lim\limits_{\substack{T\to\infty \\ \Delta_N \to 0}} \eta \min(\gamma_{31}(T), \gamma_{32}(T))/2 > 0$, and $\gamma_{31}(T), \gamma_{32}(T)$ are the smallest eigenvalues of the matrices $\sum\limits_{i \in \mathbb{N}[s^0T - \eta T, s^0 T]} Z_{1i}^\top(\hat{p}) Z_{1i}(\hat{p})/(\eta T \Delta_N)$ and $\sum\limits_{i \in \mathbb{N}[s^0T, s^0T + \eta T]} Z_{2i}^\top(\hat{p}) Z_{2i}(\hat{p})/(\eta T \Delta_N)$, respectively. Finally, for large T, from (A.19), (A.21) and (A.22), $\mathrm{SSE}_1 - \mathrm{SSE}_2 \geqslant C_1 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}>p_0\}} + C_2 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}=p_0\}}$

$$+ C_3 \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 \mathbb{I}_{\{\hat{p}<p_0\}} \geqslant \min\{C_1, C_2, C_3\} \|\boldsymbol{\mu}_1(p_0) - \boldsymbol{\mu}_2(p_0)\|^2 > 0,$$

with a positive probability. This completes the proof. □

Proposition A.6. *Suppose that Assumptions 2.1-2.4 hold. Then,*

$$\frac{1}{T} \sum_{i \in \mathbb{N}[0,\hat{s}T]} \mathcal{Z}_{1i}^\top(p_0) \mathcal{Z}_{1i}(p_0)/(t_{i+1} - t_i) - \frac{1}{T} \sum_{i \in \mathbb{N}[0,sT]} Z_{1i}^\top(p_0) Z_{1i}(p_0)/t_{i+1} - t_i) \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{L^{m/2}} \mathbf{0}, \tag{A.23}$$

$$\frac{1}{T} \sum_{i \in \mathbb{N}[\hat{s}T,T]} \mathcal{Z}_{2i}^\top(p_0) \mathcal{Z}_{2i}(p_0)/(t_{i+1} - t_i) - \frac{1}{T} \sum_{i \in \mathbb{N}[sT,T]} Z_{2i}^\top(p_0) Z_{2i}(p_0)/(t_{i+1} - t_i) \xrightarrow[\substack{T\to\infty \\ \Delta_N \to 0}]{L^{m/2}} \mathbf{0}. \tag{A.24}$$

Proof. By the definition of $Z_i(p_0)$,

$$\frac{1}{T} \sum_{i \in \mathbb{N}[0,\hat{s}T]} \frac{\mathcal{Z}_{1i}^\top(p_0) \mathcal{Z}_{1i}(p_0)}{t_{i+1} - t_i}$$

$$= \frac{1}{T}\begin{bmatrix} \sum_{i\in\mathbb{N}[0,\hat{s}T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i) & -\sum_{i\in\mathbb{N}[0,\hat{s}T]} \varphi^\top(t_i)(\ln X(t_i))(t_{i+1}-t_i) \\ -\sum_{i\in\mathbb{N}[0,\hat{s}T]} (\ln X(t_i))\varphi(t_i)(t_{i+1}-t_i) & \sum_{i\in\mathbb{N}[0,\hat{s}T]} (\ln X(t_i))^2(t_{i+1}-t_i) \end{bmatrix},$$

Note that

$$\ln X(t_i) = \begin{cases} \ln X_1(t_i), & \text{if } i \in \mathbb{N}[0,\hat{s}T] \text{ and } \hat{s} < s \\ \ln X_1(t_i)\mathbb{I}_{\{i\in\mathbb{N}[0,sT]\}} + \ln X_2(t_i)\mathbb{I}_{\{i\in\mathbb{N}[sT,\hat{s}T]\}}, & \text{if } i \in \mathbb{N}[0,\hat{s}T] \text{ and } \hat{s} > s \end{cases}.$$

First, since $\hat{s}$ is a consistent estimator of s, $\forall\, \varepsilon > 0$, $\forall\, 0 < \delta < s/2$,

$$\mathrm{P}\left(|\hat{s} - s| > \delta\right) < \varepsilon, \tag{A.25}$$

for sufficiently large T. Then,

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i) - \frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2}\right]$$

$$= \mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[sT,\hat{s}T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2} \mathbb{I}_{\{\hat{s}-s>0\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}}\right] \tag{A.26}$$

$$+\mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[\hat{s}T,sT]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2} \mathbb{I}_{\{\hat{s}-s\leqslant 0\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}}\right] \tag{A.27}$$

$$+\mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[sT,\hat{s}T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2} \mathbb{I}_{\{\hat{s}-s>0\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\delta\}}\right] \tag{A.28}$$

$$+\mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[\hat{s}T,sT]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2} \mathbb{I}_{\{\hat{s}-s\leqslant 0\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\delta\}}\right]. \tag{A.29}$$

$$(A.26) \leqslant \mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[sT,\hat{s}T]} \left\|\varphi^\top(t_i)\varphi(t_i)\right\|(t_{i+1}-t_i)\right)^{m/2} \mathbb{I}_{\{\hat{s}-s>0\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}}\right]$$

$$\leqslant (pK_\varphi)^{m/2}\mathbb{E}\left[(\hat{s}-s)^{m/2}\,\mathbb{I}_{\{\hat{s}-s>0\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}}\right] \leqslant (pK_\varphi)^{m/2}(\delta)^{m/2}.$$

Similarly, $(A.27) \leqslant (pK_\varphi)^{m/2} (\delta)^{m/2}$. Further, by Cauchy-Schwartz inequality,

$$(A.28) \leqslant \left\{ \mathbb{E}\left[\left\| \frac{1}{T} \sum_{i \in \mathbb{N}[sT, \hat{s}T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1} - t_i) \right\|^m \right] \mathbb{E}\left[\mathbb{I}_{\{\hat{s}-s>0\}} \mathbb{I}_{\{|\hat{s}-s|\geqslant\delta\}} \right] \right\}^{1/2}$$

$$\leqslant (pK_\varphi)^{m/2} \left\{ \mathbb{E}\left[(\hat{s}-s)^m \right] \mathrm{P}\left(|\hat{s}-s| \geqslant \delta \right) \right\}^{1/2} < (pK_\varphi)^{m/2} 2^{m/2} \sqrt{\varepsilon},$$

where we used the fact that $|\hat{s} - s| \leqslant 2$ a.s. Following the same techniques,

$(A.29) < (pK_\varphi)^{m/2} 2^{m/2} \sqrt{\varepsilon}$. This implies that

$$\begin{aligned} &\mathbb{E}\left[\left\| \frac{1}{T} \sum_{i \in \mathbb{N}[0, \hat{s}T]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1} - t_i) - \frac{1}{T} \sum_{i \in \mathbb{N}[0, sT]} \varphi^\top(t_i)\varphi(t_i)(t_{i+1} - t_i) \right\|^{m/2} \right] \\ &< 2(pK_\varphi)^{m/2} (\delta)^{m/2} + 2(pK_\varphi)^{m/2} 2^{m/2} \sqrt{\varepsilon} = 2(pK_\varphi)^{m/2} \left((\delta)^{m/2} + \sqrt{\varepsilon} \right). \end{aligned} \tag{A.30}$$

Second, from (A.25),

$$\mathbb{E}\left[\left\| \frac{1}{T} \sum_{i \in \mathbb{N}[0, \hat{s}T]} (\ln X(t_i))\varphi(t_i)(t_{i+1} - t_i) - \frac{1}{T} \sum_{i \in \mathbb{N}[0, sT]} (\ln X_1(t_i))\varphi(t_i)(t_{i+1} - t_i) \right\|^{m/2} \right]$$

$$= \mathbb{E}\left[\left\| \frac{1}{T} \sum_{i \in \mathbb{N}[0, \hat{s}T]} (\ln X(t_i))\varphi(t_i)(t_{i+1} - t_i) - \frac{1}{T} \sum_{i \in \mathbb{N}[0, sT]} (\ln X_1(t_i))\varphi(t_i)(t_{i+1} - t_i) \right\|^{m/2} \mathbb{I}_{\{|\hat{s}-s|>\delta\}} \right] \tag{A.31}$$

$$+ \mathbb{E}\left[\left\| \frac{1}{T} \sum_{i \in \mathbb{N}[0, \hat{s}T]} (\ln X(t_i))\varphi(t_i)(t_{i+1} - t_i) - \frac{1}{T} \sum_{i \in \mathbb{N}[0, sT]} (\ln X_1(t_i))\varphi(t_i)(t_{i+1} - t_i) \right\|^{m/2} \mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}} \right]. \tag{A.32}$$

Next,

$$(A.31) \leqslant \mathbb{E}\left[\left(\frac{1}{T} \sum_{i \in \mathbb{N}[sT, T]} \left\| (\ln X_2(t_i)\varphi(t_i) \right\| (t_{i+1} - t_i) \right)^{m/2} \mathbb{I}_{\{\hat{s}>s+\delta\}} \right]$$

$$+ \mathbb{E}\left[\left(\frac{1}{T} \sum_{i \in \mathbb{N}[0, sT]} \left\| (\ln X_1(t_i)\varphi(t_i) \right\| (t_{i+1} - t_i) \right)^{m/2} \mathbb{I}_{\{\hat{s}<s-\delta\}} \right].$$

Further, by Cauchy-Schwartz inequality,

$$\mathbb{E}\left[\left(\frac{1}{T} \sum_{i \in \mathbb{N}[sT, T]} \left\| (\ln X_2(t_i)\varphi(t_i) \right\| (t_{i+1} - t_i) \right)^{m/2} \mathbb{I}_{\{\hat{s}>s+\delta\}} \right]$$

$$\leqslant \frac{1}{T^{m/2}} \left\{ \mathbb{E}\left[\left(\sum_{i \in \mathbb{N}[sT, T]} \left\| (\ln X_2(t_i)\varphi(t_i) \right\| (t_{i+1} - t_i) \right)^m \right] \mathrm{P}\left(\hat{s} > s + \delta \right) \right\}^{1/2}$$

$$\leqslant \frac{1}{T^{m/2}}(p_0K_\varphi)^{m/2}\left\{((1-s)T)^{m-1}\left(\sum_{i\in\mathbb{N}[sT,T]}\mathbb{E}\left[|\ln X_2(t_i)|^m\right](t_{i+1}-t_i)\right)\mathrm{P}(\hat{s}>s+\delta)\right\}^{1/2}$$

$$\leqslant \left(\frac{p_0K_\varphi}{T}\right)^{m/2}((1-s)T)^{m/2}\sqrt{\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\{\mathrm{P}(\hat{s}>s+\delta)\}^{1/2}$$

$$=(p_0K_\varphi(1-s))^{m/2}\sqrt{\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\sqrt{\varepsilon}.$$

Similarly,

$$\mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}\|(\ln X_1(t_i))\varphi(t_i)\|(t_{i+1}-t_i)\right)^{m/2}\mathbb{I}_{\{\hat{s}<s-\delta\}}\right]<(p_0K_\varphi s)^{m/2}\sqrt{\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\sqrt{\varepsilon}.$$

This implies that $(A.31) < 2(p_0K_\varphi(1-s))^{m/2}\sqrt{\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\sqrt{\varepsilon}$. For (A.32),

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X(t_i))\varphi(t_i)(t_{i+1}-t_i)-\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X(t_i))\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2}\mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}}\right]$$

$$\leqslant \mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[(s-\delta)T,sT]}\|(\ln X_1(t_i)\varphi(t_i)\|(t_{i+1}-t_i)\right)^{m/2}\mathbb{I}_{\{s-\delta<\hat{s}<s\}}\right]$$

$$+\mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[sT,(s+\delta)T]}\|(\ln X_2(i)\varphi(t_i)\|(t_{i+1}-t_i)\right)^{m/2}\mathbb{I}_{\{s<\hat{s}<s+\delta\}}\right]$$

$$\leqslant 2\mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]}\|(\ln X(t_i))\varphi(t_i)\|(t_{i+1}-t_i)\right)^{m/2}\mathbb{I}_{\{s-\delta<\hat{s}<s+\delta\}}\right].$$

By Cauchy Schwartz inequality and the fact that $\mathrm{P}(s-\delta<\hat{s}<s+\delta)\leqslant 1$,

$$\mathbb{E}\left[\left\|\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X(t_i))\varphi(t_i)(t_{i+1}-t_i)-\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X(t_i))\varphi(t_i)(t_{i+1}-t_i)\right\|^{m/2}\mathbb{I}_{\{|\hat{s}-s|\leqslant\delta\}}\right]$$

$$\leqslant 2\left\{\mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]}\|(\ln X(t_i))\varphi(t_i)\|(t_{i+1}-t_i)\right)^{m}\right]\right\}^{1/2}.$$

Furthermore, by Jensen's inequality and Proposition 2.3.11,

$$\mathbb{E}\left[\left(\frac{1}{T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]}(|\ln X(t_i)|)\,\|\varphi(t_i)\|(t_{i+1}-t_i)\right)^{m}\right]$$

$$= (2\delta)^m \mathbb{E}\left[\left(\frac{1}{2\delta T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]} (|\ln X(t_i)|)\,\|\varphi(t_i)\|\,(t_{i+1}-t_i)\right)^m\right]$$

$$\leqslant (2\delta)^m \frac{1}{2\delta T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]} \mathbb{E}\left[((|\ln X(t_i)|)\,\|\varphi(t_i)\|)^m\right](t_{i+1}-t_i)$$

$$\leqslant (2\delta)^m K_\varphi^m \frac{1}{2\delta T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]} \mathbb{E}\left[(|\ln X(t_i)|)^m\right](t_{i+1}-t_i)$$

$$\leqslant (2\delta)^m K_\varphi^m \sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]\frac{1}{2\delta T}\sum_{i\in\mathbb{N}[(s-\delta)T,(s+\delta)T]}(t_{i+1}-t_i) = (2\delta)^m K_\varphi^m \sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m].$$

Since δ and ε can be arbitrary small,

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X(t_i))\varphi(t_i)(t_{i+1}-t_i) - \frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X_1(t_i))\varphi(t_i)(t_{i+1}-t_i) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} \mathbf{0}_{1\times P_0}. \quad (A.33)$$

Finally, for the last term,

$$\mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X(t_i))^2(t_{i+1}-t_i) - \frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X_1(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\right]$$

$$= \mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X(t_i))^2(t_{i+1}-t_i) - \frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X_1(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\mathbb{I}_{\{|\hat{s}-s|\geqslant\delta\}}\right] \quad (A.34)$$

$$+\mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X(t_i)))^2(t_{i+1}-t_i) - \frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X_1(t_i)^2(t_{i+1}-t_i)\right|^{m/2}\mathbb{I}_{\{|\hat{s}-s|<\delta\}}\right]. \quad (A.35)$$

Further,

$$(A.34) \leqslant \mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[sT,T]}(\ln X_2(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\mathbb{I}_{\{\hat{s}>s+\delta\}}\right] + \mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X_1(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\mathbb{I}_{\{\hat{s}<s-\delta\}}\right].$$

Then, from Jensen's inequality,

$$(A.34) \leqslant \frac{1}{T}\mathbb{E}\left[\sum_{i\in\mathbb{N}[sT,T]}|\ln X_2(t_i)|^m(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}>s+\delta\}}\right] + \frac{1}{T}\mathbb{E}\left[\sum_{i\in\mathbb{N}[0,sT]}|\ln X_2(t_i)|^m(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}<s-\delta\}}\right].$$

Then, from (2.3.16),

$$
\begin{aligned}
(A.34) \quad & \leqslant 3^{m-1}\mathbb{E}\left[\sum_{i\in\mathbb{N}[sT,T]}\left(|e^{-\alpha_2(t_i-t^*)}\ln X_0^{t^*}|^m + |r_2^{t^*}(t_i-t^*)|^m\right.\right. \\
& \left.\left. +|\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}>s+\delta\}}\right] \\
& + 3^{m-1}\mathbb{E}\left[\sum_{i\in\mathbb{N}[0,sT]}\left(|e^{-\alpha_2(t_i-t^*)}\ln X_0^{t^*}|^m\right.\right. \\
& \left.\left. +|r_2^{t^*}(t_i-t^*)|^m + |\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}<s-\delta\}}\right].
\end{aligned}
\tag{A.36}
$$

First, by Assumption 2.3,

$$
\mathbb{E}\left[\sum_{i\in\mathbb{N}[sT,T]}\left(|e^{-\alpha_2(t_i-t^*)}\ln X_0^{t^*}|^m + |r_2^{t^*}(t_i-t^*)|^m + |\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}>s+\delta\}}\right]
$$

$$
= \mathbb{E}\left[|\ln X_0^{t^*}|^m\right]\left(\sum_{i\in\mathbb{N}[sT,T]}\left(|e^{-m\alpha_2(t_i-t^*)}|\right)(t_{i+1}-t_i)\right)\mathrm{P}\{\hat{s}>s+\delta\}
$$

$$
+\left(\sum_{i\in\mathbb{N}[sT,T]}\left(|r_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\right)\mathrm{P}\{\hat{s}>s+\delta\}+\mathbb{E}\left[\sum_{i\in\mathbb{N}[sT,T]}\left(|\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}>s+\delta\}}\right].
$$

Since $\mathbb{E}\left[|\ln X_0^{t^*}|^m\right] < \infty$,

$$
\mathbb{E}\left[|\ln X_0^{t^*}|^m\right]\left(\sum_{i\in\mathbb{N}[sT,T]}\left(|e^{-m\alpha_2(t_i-t^*)}|\right)(t_{i+1}-t_i)\right)\mathrm{P}\{\hat{s}>s+\delta\} < \mathbb{E}\left[|\ln X_0^{t^*}|^m\right](1-s)T\mathrm{P}\{\hat{s}>s+\delta\}.
\tag{A.37}
$$

Further,

$$
\left(\sum_{i\in\mathbb{N}[sT,T]}\left(|r_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\right)\mathrm{P}\{\hat{s}>s+\delta\} \leqslant \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha_2}\right)^m (1-s)T\mathrm{P}\{\hat{s}>s+\delta\}.
\tag{A.38}
$$

From Cauchy-Schwartz inequality and Jensen's inequality,

$$
\begin{aligned}
& \mathbb{E}\left[\sum_{i\in\mathbb{N}[sT,T]}\left(|\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}>s+\delta\}}\right] \\
& \leqslant \left\{\mathbb{E}\left[\left(\sum_{i\in\mathbb{N}[sT,T]}|\tau_2^{t^*}(t_i-t^*)|^m(t_{i+1}-t_i)\right)^2\right]\mathrm{P}\{\hat{s}>s+\delta\}\right\}^{1/2} \\
& \leqslant (1-s)T\sigma^m\left(\frac{1}{2\alpha_2}\right)^{m/2}\sqrt{C_m\mathrm{P}\{\hat{s}>s+\delta\}}.
\end{aligned}
\tag{A.39}
$$

(A.37), (A.38), (A.39), and (A.25) imply that

$$\begin{aligned}&\mathbb{E}\left[\sum_{i\in\mathbb{N}[sT,T]}\left(|e^{-\alpha_2(t_i-t^*)}\ln X_0^{t^*}|^m+|r_2^{t^*}(t_i-t^*)|^m+|\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}>s+\delta\}}\right]\\&\leqslant\left(\mathbb{E}\left[|\ln X_0^{t^*}|^m\right]+\left(\frac{K_\mu K_\varphi+\frac{1}{2}\sigma^2}{\alpha_2}\right)^m\right)\varepsilon+\sigma^m\left(\frac{1}{2\alpha_2}\right)^{m/2}\sqrt{C_m\varepsilon}.\end{aligned}\tag{A.40}$$

Similar to the proof of (A.37), (A.38), and (A.39),

$$\begin{aligned}&\mathbb{E}\left[\sum_{i\in\mathbb{N}[0,sT]}\left(|e^{-\alpha_2(t_i-t^*)}\ln X_0^{t^*}|^m+|r_2^{t^*}(t_i-t^*)|^m+|\tau_2^{t^*}(t_i-t^*)|^m\right)(t_{i+1}-t_i)\mathbb{I}_{\{\hat{s}<s-\delta\}}\right]\\&\leqslant\left(\mathbb{E}\left[|\ln X_0^{t^*}|^m\right]+\left(\frac{K_\mu K_\varphi+\frac{1}{2}\sigma^2}{\alpha_2}\right)^m\right)\varepsilon+\sigma^m\left(\frac{1}{2\alpha_2}\right)^{m/2}\sqrt{C_m\varepsilon}.\end{aligned}\tag{A.41}$$

(A.36), (A.40) and (A.41) imply that

$$(A.34)\leqslant 2\left(\left(\mathbb{E}\left[|\ln X_0^{t^*}|^m\right]+\left(\frac{K_\mu K_\varphi+\frac{1}{2}\sigma^2}{\alpha_2}\right)^m\right)\varepsilon+\sigma^m\left(\frac{1}{2\alpha_2}\right)^{m/2}\sqrt{C_m\varepsilon}\right).\tag{A.42}$$

Following the same technique,

$$\begin{aligned}(A.35)\leqslant&\ \mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[(s-\delta)T,sT]}(\ln X_1(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\mathbb{I}_{\{s-\delta<\hat{s}<s\}}\right]\\&+\mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[sT,(s+\delta)T]}(\ln X_2(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\mathbb{I}_{\{s<\hat{s}<s+\delta\}}\right]\\\leqslant&\ \mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[(s-\delta)T,sT]}(\ln X_1(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\right]\\&+\mathbb{E}\left[\left|\frac{1}{T}\sum_{i\in\mathbb{N}[sT,(s+\delta)T]}(\ln X_2(t_i))^2(t_{i+1}-t_i)\right|^{m/2}\right].\end{aligned}$$

From Jensen's inequality, this gives

$$\begin{aligned}(A.35)\leqslant&(\delta)^{m/2}\frac{1}{\delta T}\sum_{i\in\mathbb{N}[(s-\delta)T,sT]}\mathbb{E}\left[|\ln X_1(t_i)|^m\right](t_{i+1}-t_i)+(\delta)^{m/2}\frac{1}{\delta T}\\&\sum_{i\in\mathbb{N}[sT,(s+\delta)T]}\mathbb{E}\left[|\ln X_2(t_i)|^m\right](t_{i+1}-t_i).\end{aligned}$$

Proposition 2.3.11 gives that

$$(A.35)\leqslant\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]\,(\delta)^m\,.\tag{A.43}$$

Since δ and ε can be arbitrary small, together with (A.42) and (A.43),

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\ln X_i)^2(t_{i+1}-t_i)-\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}(\ln X_i)^2(t_{i+1}-t_i)\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} 0. \tag{A.44}$$

$(A.30),(A.33)$ and $(A.44)$ imply that

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\mathcal{Z}_{1i}^{\top}(p_0)\mathcal{Z}_{1i}(p_0)/(t_{i+1}-t_i)-\frac{1}{T}\sum_{i\in\mathbb{N}[0,sT]}Z_{1i}^{\top}(p_0)Z_{1i}(p_0)/(t_{i+1}-t_i)\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}}\mathbf{0}.$$

By using similar techniques, one proves the second statement. This completes the proof.

□

Proposition A.7. *Suppose that Assumptions 2.1-2.4 hold. Then,*

$$\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\frac{\mathcal{Z}_{1i}^{\top}(p_0)\mathcal{Z}_{1i}(p_0)}{t_{i+1}-t_i}\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}} s\mathbf{\Sigma}_1, \quad \textit{and} \quad \frac{1}{T}\sum_{i\in\mathbb{N}[\hat{s}T,T]}\frac{\mathcal{Z}_{2i}^{\top}(p_0)\mathcal{Z}_{2i}(p_0)}{t_{i+1}-t_i}\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}}(1-s)\mathbf{\Sigma}_2. \tag{A.45}$$

Proof. To prove the first statement, it suffices to combines Proposition A.6, Proposition 2.3.17 along with some algebraic computations. The proof of the second statement in (A.45) is similar. This completes the proof. □

Proposition A.8. *If Assumptions 2.1-2.4 hold, then,*

$$\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\epsilon_i\mathcal{Z}_{1i}(p_0)/(t_{i+1}-t_i)-\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,sT]}\epsilon_iZ_{1i}(p_0)/(t_{i+1}-t_i)\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}}\vec{0}_{1\times(p_0+1)}; \tag{A.46}$$

$$\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[\hat{s}T,T]}\epsilon_i\mathcal{Z}_{2i}(p_0)/(t_{i+1}-t_i)-\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[sT,T]}\epsilon_iZ_{2i}(p_0)/(t_{i+1}-t_i)\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^{m/2}}\vec{0}_{1\times(p_0+1)}. \tag{A.47}$$

Proof. To prove (A.46), the following inequality is used.

$$\begin{aligned}&\left\|\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\epsilon_i\mathcal{Z}_{1i}(p_0)/(t_{i+1}-t_i)-\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,sT]}\epsilon_i Z_{1i}(p_0)/(t_{i+1}-t_i)\right\|^{m/2}\\
&\leqslant 3^{m/2-1}\left(\left\|\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\epsilon_i\mathcal{Z}_{1i}(p_0)/(t_{i+1}-t_i)-\frac{\sigma}{\sqrt{T}}\int_0^{\hat{s}T}(\varphi(t),-\ln X(t))\,dB_t\right\|^{m/2}\right.\\
&+\left\|\frac{\sigma}{\sqrt{T}}\int_0^{\hat{s}T}(\varphi(t),-\ln X(t))\,dB_t-\frac{\sigma}{\sqrt{T}}\int_0^{sT}(\varphi(t),-\ln X(t))\,dB_t\right\|^{m/2}\\
&\left.+\left\|\frac{\sigma}{\sqrt{T}}\int_0^{sT}(\varphi(t),-\ln X(t))\,dB_t-\frac{1}{\sqrt{T}}\sum_{i\in\mathbb{N}[0,sT]}\epsilon_i Z_{1i}(p_0)/(t_{i+1}-t_i)\right\|^{m/2}\right).\end{aligned}$$

One can prove that,

$$\sum_{i\in\mathbb{N}[0,\hat{s}T]}\epsilon_i\mathcal{Z}_{1i}(p_0)/((t_{i+1}-t_i)\sqrt{T})-\sigma\int_0^{\hat{s}T}(\varphi(t),-\ln X(t))\,dB_t/\sqrt{T}\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^m}\mathbf{0}_{1\times(p+1)},$$

$$\sum_{i\in\mathbb{N}[0,sT]}\epsilon_i Z_{1i}(p_0)/((t_{i+1}-t_i)\sqrt{T})-\sigma\int_0^{sT}(\varphi(t),-\ln X(t))\,dB_t/\sqrt{T}\xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{L^m}\mathbf{0}.$$

By Lemma A.6-Lemma A.7,

$$\sigma\int_0^{\hat{s}T}(\varphi(t),-\ln X(t))\,dB_t/\sqrt{T}-\sigma\int_0^{sT}(\varphi(t),-\ln X(t))\,dB_t/\sqrt{T}\xrightarrow[T\to\infty]{L^{m/2}}\mathbf{0}.$$

This completes the proof of (A.46), and (A.47) is proven by similar techniques. □

Proposition A.9. *If Assumption 2.1-2.3 and Assumption 2.4 hold, then, for some* $0 < a^* < \boldsymbol{a}/2$*,*

$$\left(\left\|\sum_{i\in\mathbb{N}[0,\hat{s}T]}\varepsilon_i\mathcal{Z}_{1i}(p)/(t_{i+1}-t_i)\right\|,\left\|\sum_{i\in\mathbb{N}[\hat{s}T,T]}\varepsilon_i\mathcal{Z}_{2i}(p)/(t_{i+1}-t_i)\right\|\right)=O_p(\sqrt{T}\log^{a^*}(T)).\tag{A.48}$$

Proof. The proof follows from the triangle inequality and Proposition A.5. □

Proposition A.6-Proposition A.9 imply that $\hat{\theta}_T(\hat{s})$ obtained from discretized version is consistent.

Proof of Proposition 2.4.8. From the log-likelihood function defined in (2.3.17) and

SIC information criterion in (2.4.11), along with the fact that, by (2.4.3),

$$Y_i = Z_{ki}(p_0)\boldsymbol{\mu}_k(p_0) + \epsilon_i, \quad k = 1, 2.$$

Then,

$$\begin{aligned}\mathrm{IC}(1,p) =& -2\left(\frac{1}{2\sigma^2}\left(\sum_{i\in\mathbb{N}[0,T]} Y_i^2/(t_{i+1}-t_i) - \sum_{i\in\mathbb{N}[0,\hat{s}T]} \frac{(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_1(p_0)+\varepsilon_i-\mathcal{Z}_{1i}(p)\hat{\boldsymbol{\mu}}_1(p))^2}{(t_{i+1}-t_i)}\right.\right.\\ &\left.\left. -\sum_{i\in\mathbb{N}[\hat{s}T,T]} \frac{(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0)+\varepsilon_i-\mathcal{Z}_{2i}(p)\hat{\boldsymbol{\mu}}_2(p))^2}{(t_{i+1}-t_i}\right)\right) + 2(p+1)\log(N),\end{aligned}$$

and

$$\begin{aligned}\mathrm{IC}(0,p) = -2\left(\frac{1}{2\sigma^2}\left(\sum_{i\in\mathbb{N}[0,T]} \frac{Y_i^2}{t_{i+1}-t_i} - \sum_{i\in\mathbb{N}[0,T]} \frac{(Z_{1i}(p_0)\boldsymbol{\mu}_1(p_0)+\varepsilon_i-Z_{1i}(p)\hat{\boldsymbol{\mu}}_1(p))^2}{t_{i+1}-t_i}\right)\right)\\ +(p+1)\log(N).\end{aligned}$$

First, the value of $\mathrm{IC}(0, p_0)$, $\mathrm{IC}(c, p)$, for $c = 1$, or $p \neq p_0$ is compared under the condition $c^0 = 0$.

(1) $c = 1, p = p_0$: In this case, $\boldsymbol{\mu}_1(p_0) = \boldsymbol{\mu}_2(p_0) = \boldsymbol{\mu}(p_0)$ and one verifies that

$$\begin{aligned}\mathrm{IC}(0,p_0) - \mathrm{IC}(1,p_0) =& \frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,T]} \frac{\left(Z_i(p_0)\boldsymbol{\theta}(p_0) - Z_i(p_0)\hat{\theta}(p_0)\right)^2}{t_{i+1}-t_i}\\ &-\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,\hat{s}T]} \frac{(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_1(p_0)-\mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_1(p_0))^2}{t_{i+1}-t_i}\\ &-\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[\hat{s}T,T]} \frac{(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0)-\mathcal{Z}_{2i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0))^2}{t_{i+1}-t_i} - (p_0+1)\log(N)\\ &+\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,T]} \frac{2\varepsilon_i\left(Z_i(p_0)\boldsymbol{\theta}(p_0) - Z_i(p_0)\hat{\theta}(p_0)\right)}{t_{i+1}-t_i}\\ &-\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,\hat{s}T]} \frac{2\varepsilon_i\,(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_1(p_0)-\mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_1(p_0))}{t_{i+1}-t_i}\\ &-\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[\hat{s}T,T]} \frac{2\varepsilon_i\,(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0)-\mathcal{Z}_{2i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0))}{t_{i+1}-t_i}.\end{aligned}$$

From Proposition 2.3.18, $\hat{\theta}(p_0) = \boldsymbol{\theta}(p_0) + \sigma Q^{-1}(s, T, p_0)W(s, T, p_0)$,

$$\frac{1}{\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\left(Z_i(p_0)\boldsymbol{\theta}(p_0)-Z_i(p_0)\hat{\theta}(p_0)\right)^2/(t_{i+1}-t_i)$$

$$=\frac{1}{T}(TQ^{-1}(s,T,p_0))\frac{1}{\sqrt{T}}W(s,T,p_0))^\top\frac{1}{T}\sum_{i\in\mathbb{N}[0,T]}\frac{Z_i^\top(p_0)Z_i(p_0)}{t_{i+1}-t_i}\times$$

$$(TQ^{-1}(s,T,p_0))\frac{1}{\sqrt{T}}W(s,T,p_0)).$$

By combining Proposition 2.3.17, Proposition A.4, Proposition A.6, Proposition A.7 and Proposition A.18,

$$\sum_{i\in\mathbb{N}[0,T]}\left(Z_i(p_0)\boldsymbol{\theta}(p_0)-Z_i(p_0)\hat{\theta}(p_0)\right)^2/(t_{i+1}-t_i)=O_p(\log^{2a^*}(T)),\tag{A.49}$$

$$\sum_{i\in\mathbb{N}[0,\hat{s}T]}(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_2(p_0)-\mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0))^2/(t_{i+1}-t_i)=O_p(\log^{2a^*}(T)),$$

$$\sum_{i\in\mathbb{N}[\hat{s}T,T]}(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0)-\mathcal{Z}_{2i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0))^2/(t_{i+1}-t_i)=O_p(\log^{2a^*}(T)).$$

Further, from Proposition A.4 and Proposition A.9,

$$\sum_{i\in\mathbb{N}[0,T]}2\varepsilon_i\left(Z_i(p_0)\boldsymbol{\theta}(p_0)-Z_i(p_0)\hat{\theta}(p_0)\right)/(t_{i+1}-t_i)=O_p(\log^{2a^*}(T)),\tag{A.50}$$

$$\sum_{i\in\mathbb{N}[0,\hat{s}T]}2\varepsilon_i\left(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_2(p_0)-\mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0)\right)/(t_{i+1}-t_i)=O_p(\log^{2a^*}(T)),$$

$$\sum_{i\in\mathbb{N}[\hat{s}T,T]}2\varepsilon_i\left(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0)-\mathcal{Z}_{2i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0)\right)/(t_{i+1}-t_i)=O_p(\log^{2a^*}(T)).$$

Then, together with Assumption 2.4, $(p_0+1)\log(N)=O(\log^a(T))$,

$\mathrm{IC}(0,p_0)-\mathrm{IC}(1,p_0)<0$, whenever T tends to infinity and Δ_N tends to 0 i.e.

$\lim\limits_{\substack{T\to\infty\\ \Delta_N\to 0}}\mathrm{P}\left(\mathrm{IC}(0,p_0)-\mathrm{IC}(1,p_0)>0\right)=0.$

(2)-(5) Similarly, it is proven that $\lim\limits_{\substack{T\to\infty\\ \Delta_N\to 0}}\mathrm{P}\left(\mathrm{IC}(0,p_0)-\mathrm{IC}(1,p_0)>0\right)=1$ for the cases where **(2)** $c=1,p=p^*>p_0$; **(3)** $c=1,p=p_*<p_0$; **(4)** $c=0,p=p^*>p_0$ and **(5)** $c=0,p=p_*<p_0$.

This completes the proof of first part.

Second, suppose that $c^0 = 1$, the value of $\mathrm{IC}(1, p_0)$, $\mathrm{IC}(c, p)$, for $c = 0$, or $p \neq p_0$ is compared.

(1). $c = 0, p = p_0$:

$$\frac{\mathrm{IC}(0,p_0) - \mathrm{IC}(1,p_0)}{T} = \frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{\left(Z_i(p_0)\boldsymbol{\mu}_1(p_0) - Z_i(p_0)\hat{\theta}(p_0)\right)^2}{t_{i+1}-t_i}$$

$$-\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\frac{\mathcal{Z}_{1i}(p_0)\left(\boldsymbol{\mu}_1(p_0) - \hat{\boldsymbol{\mu}}_1(p_0)\right)^2}{t_{i+1}-t_i}$$

$$-\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[\hat{s}T,T]}\frac{\mathcal{Z}_{2i}(p_0)\left(\boldsymbol{\mu}_2(p_0) - \hat{\boldsymbol{\mu}}_2(p_0)\right)^2}{t_{i+1}-t_i} - \frac{(p_0+1)\log(N)}{T}.$$

$$+\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,T]}\frac{2\varepsilon_i Z_i(p_0)\left(\boldsymbol{\theta}(p_0) - \hat{\theta}(p_0)\right)}{t_{i+1}-t_i}$$

$$-\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\frac{2\varepsilon_i\left(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_1(p_0) - \mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_1(p_0)\right)}{t_{i+1}-t_i}$$

$$-\frac{1}{T\sigma^2}\sum_{i\in\mathbb{N}[\hat{s}T,T]}\frac{2\varepsilon_i\left(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0) - \mathcal{Z}_{2i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0)\right)}{t_{i+1}-t_i}.$$

From some algebraic computations, one can proves that

$$\sum_{i\in\mathbb{N}[0,T]}\left(Z_i(p_0)\boldsymbol{\theta}(p_0) - Z_i(p_0)\hat{\theta}(p_0)\right)^2/((t_{i+1}-t_i)T\sigma^2) \geqslant C_0\|\boldsymbol{\mu}_1 - \boldsymbol{\mu}_2\| > 0$$

for some positive constant C_0. Further, since $\hat{\boldsymbol{\mu}}_1(p_0) - \boldsymbol{\mu}_1(p_0) = \sigma Q_T(\hat{s})W_T(\hat{s})$

$$\sum_{i\in\mathbb{N}[0,\hat{s}T]}\left(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_1(p_0) - \mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_1(p_0)\right)^2/((t_{i+1}-t_i)T\sigma^2)$$

$$= \frac{1}{T}(TQ_T^{-1}(\hat{s},p_0)\frac{1}{\sqrt{T}}W_T(\hat{s},p_0))^\top\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]}\frac{\mathcal{Z}_{1i}^\top(p_0)\mathcal{Z}_{1i}(p_0)}{t_{i+1}-t_i}\times$$

$$(TQ_T^{-1}(\hat{s},p_0)\frac{1}{\sqrt{T}}W_T(\hat{s},p_0)).$$

By Proposition A.6, Proposition A.7 and Proposition A.18, $\left\|\sum_{i\in\mathbb{N}[0,\hat{s}T]}\mathcal{Z}_{1i}^\top(p_0)\mathcal{Z}_{1i}(p_0)/((t_{i+1}-t_i)T)\right\| = O_P(1)$, $\|TQ_T^{-1}(\hat{s},p_0)\| = O_P(1)$, and then, together with Proposition A.4,

$$\|W_T(s,p)\|/T = o_p(1), \quad \|W_T(1,p) - W_T(s,p)\|/T = o_p(1),$$

$$\sum_{i\in\mathbb{N}[0,\hat{s}T]} \left(\mathcal{Z}_{1i}(p_0)\boldsymbol{\mu}_1(p_0) - \mathcal{Z}_{1i}(p_0)\hat{\boldsymbol{\mu}}_1(p_0)\right)^2/(t_{i+1}-t_i) = O_p(\log^{a^*}(T)),$$

$$\sum_{i\in\mathbb{N}[\hat{s}T,T]} \left(\mathcal{Z}_{2i}(p_0)\boldsymbol{\mu}_2(p_0) - \mathcal{Z}_{2i}(p_0)\hat{\boldsymbol{\mu}}_2(p_0)\right)^2/(t_{i+1}-t_i) = O_p(\log^{a^*}(T)).$$

Hence, by using Proposition A.9,

$$\sum_{i\in\mathbb{N}[0,T]} 2\varepsilon_i Z_i(p_0)\left(\boldsymbol{\theta}(p_0) - \hat{\theta}(p_0)\right)/((t_{i+1}-t_i)T) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0,$$

$$\left(\frac{1}{T}\sum_{i\in\mathbb{N}[0,\hat{s}T]} \frac{2\varepsilon_i\mathcal{Z}_{1i}(p_0)(\boldsymbol{\mu}_1(p_0)-\hat{\boldsymbol{\mu}}_1(p_0))}{t_{i+1}-t_i}, \frac{1}{T}\sum_{i\in\mathbb{N}[\hat{s}T,T]} \frac{2\varepsilon_i\mathcal{Z}_{2i}(p_0)(\boldsymbol{\mu}_2(p_0)-\hat{\boldsymbol{\mu}}_2(p_0))}{t_{i+1}-t_i}\right) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} \mathbf{0}.$$

Therefore, by Assumption 2.4, $\mathrm{IC}(0,p_0) - \mathrm{IC}(1,p_0)/T > 0$, whenever T is large and Δ_N is arbitrary small, i.e. $\lim_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}\left(\mathrm{IC}(0,p_0) - \mathrm{IC}(1,p_0) > 0\right) = 1$.

(2)-(5). Similarly, it is proven that $\lim_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}\left(\mathrm{IC}(0,p_0) - \mathrm{IC}(1,p_0) > 0\right) = 1$ for the cases where **(2)** $c = 0, p = p^* > p_0$; **(3)** $c = 0, p = p_* < p_0$; **(4)** $c = 1, p = p^* > p_0$ and **(5)** $c = 1, p = p_* < p_0$

Further, for the second claim, by (2.4.15) and definition of $\hat{p}$ and $\hat{c}$,

$$\lim_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(c^0,p_0) \le \mathrm{IC}(\hat{c},\hat{p})) = 1, \text{ and } \lim_{\substack{T\to\infty\\ \Delta_N\to 0}} \mathrm{P}(\mathrm{IC}(\hat{c},\hat{p}) \le \mathrm{IC}(c^0,p_0)) = 1.$$

This implies

$$\mathrm{IC}(\hat{c},\hat{p}) - \mathrm{IC}(c^0,p_0) \xrightarrow[\substack{T\to\infty\\ \Delta_N\to 0}]{\mathrm{P}} 0.$$

Set $\hat{p} = \hat{p}_{(T)}$ and suppose $\hat{p} - p_0 \nrightarrow 0$ in probability, as $T \to \infty$. i.e. $\exists\, \epsilon_0 > 0, \forall\, T > 0, \exists\, T_0 > T$, such that $|\hat{p}_{(T_0)} - p_0| > \epsilon_0$, which implies that $\hat{p}_{(T_0)} > p_0 + \epsilon_0$ or $\hat{p}_{(T_0)} < p_0 - \epsilon_0$ hold. For $\hat{p}_{(T_0)} > p_0 + \epsilon_0$, there exists some $\eta_0 > 0$, such that $\hat{p}_{(T_0)} = p_0 + \epsilon_0 + \eta_0 > p_0 + \epsilon_0$, so $\mathrm{IC}(\hat{c},\hat{p}_{T_0}) = \mathrm{IC}(\hat{c},p_0+\epsilon_0+\eta_0)$. This is a contradiction with the known condition: $\mathrm{IC}(\hat{c}, \hat{p}_{T_0}) - \mathrm{IC}(c^0, p_0) \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$, since from Proposition 2.4.8, $\mathrm{IC}(c^0, p_0) < \mathrm{IC}(\hat{c}, p_0 + \epsilon_0 + \eta_0)$. For $\hat{p}_{(T_0)} < p_0 - \epsilon_0$, there exists some $\eta_0 > 0$, such that $\hat{p}_{(T_0)} = p_0 - \epsilon_0 - \eta_0 < p_0 - \epsilon_0$, so $\mathrm{IC}(\hat{c}, \hat{p}_{(T_0)}) = \mathrm{IC}(\hat{c}, p_0 - \epsilon_0 - \eta_0)$. This is a contradiction with the known condition: $\mathrm{IC}(\hat{c}, \hat{p}_{(T_0)}) - \mathrm{IC}(c^0, p_0) \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$, since from Proposition 2.4.8, $\mathrm{IC}(c^0, p_0) < \mathrm{IC}(\hat{c}, p_0 - \epsilon_0 - \eta_0)$. So, $\hat{p} - p_0 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$. Similarly, using the same technique, $\hat{c} - c^0 \xrightarrow[\substack{T\to\infty \\ \Delta_N\to 0}]{\mathrm{P}} 0$. This completes the proof. □

The following proposition plays important role in proving that $\hat{s}$ is a consistent estimator of the parameter s.

Proposition A.10. *Let $t^* = sT$. There exists an $L_0 > 0$, such that for all $L > L_0$, the minimum eigenvalues of the matrices $\sum_{i\in\mathbb{N}[t^*,t^*+L]} \frac{Z_{2i}^\top(p_0)Z_{2i}(p_0)}{((t_{i+1}-t_i)L)}$ and of $\sum_{i\in\mathbb{N}[t^*-L,t^*]} \frac{Z_{1i}^\top(p_0)Z_{1i}(p_0)}{((t_{i+1}-t_i)L)}$ and their respective continuous time versions $\frac{1}{L}Q_{[t^*,t^*+L]}(p_0)$ and $\frac{1}{L}Q_{[t^*-L,t^*]}(p_0)$ are all bounded away from 0.*

The proof follows from Lemma A.4 and Lemma A.5 in Appendix A.2. Note that, Proposition A.10 serves to weaken the assumption in Chen et al. [2020], Chen et al. [2018]. More precisely, the established result shows that the assumptions in the quoted paper are stronger than needed.

Proposition A.11. $\psi(s, c^0, p_0) \sim \chi_q^2(\tilde{\Delta})$ and $\psi_0(s, c^0, p_0) \sim \chi_q^2$.

Proof. The proof follows from Theorem 5.1.3 in Mathai and Provost [1992]. Since that

$$\tilde{\varsigma} \sim \mathcal{N}_{(c^0+1)(p_0+1)}\left(\tilde{G}^*(c^0, p_0)\tilde{r}_0, \sigma^2\tilde{G}^*(c^0, p_0)\tilde{M}(c^0, p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right),$$

$$\tilde{\varsigma}_0 \sim \mathcal{N}_{(c^0+1)(p_0+1)}\left(0, \sigma^2\tilde{G}^*(c^0, p_0)\tilde{M}(c^0, p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right),$$

$\psi(s, c^0, p_0) = \tilde{\varsigma}^\top\Gamma\tilde{\varsigma}$ and $\psi_0(s, c^0, p_0) = \tilde{\varsigma}_0^\top\Gamma\tilde{\varsigma}_0$. To prove that $\psi(s, c^0, p_0) = \tilde{\varsigma}^\top\Gamma(c^0, p_0)\tilde{\varsigma} \sim \chi_q^2(\tilde{\Delta})$ *and* $\psi_0(s, c^0, p_0) = \tilde{\varsigma}_0^\top\Gamma(c^0, p_0)\tilde{\varsigma}_0 \sim \chi_q^2$, it suffices to apply Theorem 5.1.3 in

Mathai and Provost [1992]. As a consequence, the following four equalities need to be proven.

(1) $\mathrm{trace}(\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0))=q$,

$(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(c^0,p_0)(\tilde{G}^*(c^0,p_0)\tilde{r}_0)=\Delta$,

(2) $\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma(c^0,p_0)\times$
$\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)=\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma(c^0,p_0)\sigma^2\times$
$\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)$,

(3) $(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma(c^0,p_0)(\tilde{G}^*(c^0,p_0)\tilde{r}_0)$
$=(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(c^0,p_0)(\tilde{G}^*(c^0,p_0)\tilde{r}_0)$,

(4) $(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top(\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0))^2$
$=(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)$.

(1) First, since $\tilde{G}^*(c^0,p_0)=\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}$, and $\Gamma(c^0,p_0)=\dfrac{1}{\sigma^2}\tilde{M}^\top(c^0,p_0)\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)$,
$\mathrm{trace}\left(\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)=\mathrm{trace}\left(\frac{1}{\sigma^2}\tilde{M}^\top(c^0,p_0)\left(\tilde{M}(c^0,p_0)\times\right.\right.$
$\left.\left.\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)$

$$=\mathrm{trace}\left(\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)=q,$$

where the property $\mathrm{trace}(A^\top B)=\mathrm{trace}(BA^\top)$ is used. Then,

$$(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(c^0,p_0)(\tilde{G}^*(c^0,p_0)\tilde{r}_0)=\left(\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\times\right.\right.$$

$$\left.\left.\tilde{M}^\top(c^0,p_0)\right)^{-1}r_0\right)^\top\frac{1}{\sigma^2}\tilde{M}^\top(c^0,p_0)\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\times$$

$$\tilde{M}(c^0,p_0)\left(\tilde{\Sigma}_{c^0}^{-1}(p_0)M^\top\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}r_0\right)$$

$$=\frac{1}{\sigma^2}r_0^\top\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\times$$

$$\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)$$

$$\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1} r_0$$

$$= \frac{1}{\sigma^2} r_0^\top \left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1} r_0 = \Delta.$$

(2) For assertion (2),

$$\begin{aligned}
&\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma\sigma^2\tilde{G}^*(c^0,p_0)\times\\
&\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\\
&= \sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\left(\tfrac{1}{\sigma^2}M^\top\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\right)\\
&\times\left(\sigma^2\tilde{\Sigma}_{c^0}^{-1}(p_0)M^\top\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}M\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\\
&\times\left(\tfrac{1}{\sigma^2}M^\top\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\right)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\\
&= \sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\left(\tfrac{1}{\sigma^2}M^\top\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}M\right)\times\\
&\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\\
&= \sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0).
\end{aligned}$$

(3) From part (ii), $\Gamma\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma = \Gamma$. Then,

$$\begin{aligned}
&(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma(\tilde{G}^*(c^0,p_0)\tilde{r}_0)\\
&\qquad = (\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(\tilde{G}^*(c^0,p_0)\tilde{r}_0).
\end{aligned}$$

(4) Since $\Gamma\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\Gamma = \Gamma$,

$$\begin{aligned}
&(\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top(\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0))^2\\
&= (\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top(\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0))(\Gamma(c^0,p_0)\times\\
&\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0))\\
&= (\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Gamma(c^0,p_0)\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0).
\end{aligned}$$

It is a special case when $r_0 = 0$. This completes the proof. □

Proof of Proposition 2.4.9. Let x be a continuous point of the cdf of $\psi(s,p_0)$,

$$\lim_{T\to\infty} F_{\tilde{\psi}_T}(x) = \lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x\right)$$

$$= \lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} = c^0, \hat{p} = p_0\right) + \lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} = c^0, \hat{p} \neq p_0\right)$$

$$+ \lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} \neq c^0, \hat{p} = p_0\right) + \lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} \neq c^0, \hat{p} \neq p_0\right).$$

Since

$$\lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} = c^0, \hat{p} \neq p_0\right) \leqslant \lim_{T\to\infty} \mathrm{P}\left(\hat{p} \neq p_0\right) = 0,$$

$$\lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} \neq c^0, \hat{p} = p_0\right) \leqslant \lim_{T\to\infty} \mathrm{P}\left(\hat{c} \neq c^0\right) = 0,$$

$$\lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} \neq c^0, \hat{p} \neq p_0\right) \leqslant \lim_{T\to\infty} \mathrm{P}\left(\hat{c} \neq c^0\right) = 0,$$

then, together with (2.4.14),

$$\lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x\right) = \lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x, \hat{c} = c^0, \hat{p} = p_0\right).$$

Then,

$$\lim_{T\to\infty} \mathrm{P}\left(\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \leqslant x\right) = \lim_{T\to\infty} \mathrm{P}\left(\psi_T(\hat{s},c^0,p_0) \leqslant x\right) = \mathrm{P}\left(\psi(s,c^0,p_0) \leqslant x\right).$$

Which implies that $\lim_{T\to\infty} F_{\tilde{\psi}_T}(x) = F_{\psi}(x)$. This is $\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \xrightarrow[T\to\infty]{D} \psi(s,c^0,p_0)$. Since $\psi(s,c^0,p_0) = \xi^\top\Gamma\xi$ and $\xi \sim \mathcal{N}\left(G^*(c^0,p_0)r_0, \sigma^2 G^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)$, by Proposition A.11, if $r_0 \neq 0$, $\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \xrightarrow[T\to\infty]{\mathrm{D}} \psi(s,c^0,p_0) \sim \chi_q^2(\tilde{\Delta})$. If $r_0 = 0$, then, $\tilde{\psi}_T(\hat{s},\hat{c},\hat{p}) \xrightarrow[T\to\infty]{\mathrm{D}} \psi(s,c^0,p_0) \sim \chi_q^2$. This completes the proof. □

A.2 On the solution of GEOU process and other important results

Proposition A.12. *Suppose that $\sigma(t)$ is a nonrandom square integrable function on $[0,T]$. Then the Itô's integral*

$$\mathcal{I}(t) = \int_0^t \sigma(u)dB_u \tag{A.51}$$

is a mean zero Gaussian process with $\mathrm{Var}(\mathcal{I}(t)) = \int_0^t \sigma^2(u)du$.

Proof. It is easy to check that $\mathcal{I}(t)$ is a martingale, thus, $\mathbb{E}[\mathcal{I}(t)] = \mathbb{E}[\mathcal{I}(0)] = 0$, and then, $\mathrm{Var}(\mathcal{I}(t)) = \int_0^t \sigma^2(u)du$. To show the normal distribution of $\mathcal{I}(t)$, for any fixed

$\lambda \in \mathbb{R}$, define

$$dU_t = \int_0^t \lambda\sigma(u)dB_u - \frac{1}{2}\int_0^t (\lambda\sigma(u))^2 du$$

It is well known that $U(t)$ is a generalized geometric Brownian motion and $(e^{U_t}, t \geq 0)$ is a martingale. Thus has constant mean value

$$\mathbb{E}[e^{U_t}] = \mathbb{E}[e^{U_0}] = e^0 = 1$$

which is the same as

$$\mathbb{E}\left[\exp\left\{\int_0^t \lambda\sigma(u)dB_u - \frac{1}{2}\int_0^t (\lambda\sigma(u))^2 du\right\}\right] = 1.$$

The second integral is not a random variable, and

$$\mathbb{E}\left[e^{\lambda\mathcal{I}(t)}\right] = \exp\left\{\frac{1}{2}\lambda^2\int_0^t (\sigma(u))^2 du\right\}.$$

is a function of λ, the left hand side is the moment generating function of $\mathcal{I}(t)$, while the right hand side is the mgf of a normal distribution. Further, it is proven that for all $0 < t_1 < t_2 < \cdots < t_n < T$, the finite multivariate random variable $(\mathcal{I}(t_1), \mathcal{I}(t_2), \cdots, \mathcal{I}(t_n))$ follows a multivariate normal distribution. To make use of the independet increament of Brownian motion, let

$$\begin{bmatrix} \mathcal{I}(t_1) \\ \mathcal{I}(t_2 - t_1) \\ \mathcal{I}(t_3 - t_2) \\ \vdots \\ \mathcal{I}(t_n - t_{n-1}) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \begin{bmatrix} \mathcal{I}(t_1) \\ \mathcal{I}(t_2) \\ \mathcal{I}(t_3) \\ \vdots \\ \mathcal{I}(t_n) \end{bmatrix} = \mathbf{M} \begin{bmatrix} \mathcal{I}(t_1) \\ \mathcal{I}(t_2) \\ \mathcal{I}(t_3) \\ \vdots \\ \mathcal{I}(t_n) \end{bmatrix},$$

and each element of $(\mathcal{I}(t_1), \mathcal{I}(t_2 - t_1), \cdots, \mathcal{I}(t_n - t_{n-1}))$ is normal distributed and they are mutually independent, which implies that $(\mathcal{I}(t_1), \mathcal{I}(t_2 - t_1), \cdots, \mathcal{I}(t_n - t_{n-1}))$ follows a multivariate normal with mean zero. In addition, the matrix $\mathbf{M}$ is invertable, then, $(\mathcal{I}(t_1), \mathcal{I}(t_2), \cdots, \mathcal{I}(t_n))^\top = \mathbf{M}^{-1}(\mathcal{I}(t_1), \mathcal{I}(t_2 - t_1), \cdots, \mathcal{I}(t_n - t_{n-1}))^\top$, which implies that the finite multivariate random variable $(\mathcal{I}(t_1), \mathcal{I}(t_2), \cdots, \mathcal{I}(t_n))$ follows multivariate normal. This finishes the proof that $\mathcal{I}(t)$ is a mean zero Gaussian process. □

Proof of Proposition 2.2.2. **(1)**. By (2.2.4),

$$\mathbb{E}\left[|\ln X(t)|^m\right] = \mathbb{E}\left[\left|e^{-\alpha t}\ln X_0 + e^{-\alpha t}\int_0^t e^{\alpha s}(L(s) - \frac{1}{2}\sigma^2)ds + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right|^m\right]$$

$$\leqslant 3^{m-1}\left(e^{-m\alpha t}\mathbb{E}\left[|\ln X_0|^m\right] + \left|e^{-\alpha t}\int_0^t e^{\alpha s}(L(s) - \frac{1}{2}\sigma^2)ds\right|^m + \sigma^m e^{-m\alpha t}\mathbb{E}\left[\left|\int_0^t e^{\alpha s}dB_s\right|^m\right]\right).$$

From Assumption 2.3,

$$\begin{cases} -K_\mu K_\varphi \leqslant L(t) \leqslant K_\mu K_\varphi \\ \sum_{i=1}^p |\mu_i| \leqslant K_\mu \end{cases} \tag{A.52}$$

for some positive constant $K_\varphi > 0$, $K_\mu > 0$,

$$\begin{aligned} &\left|e^{-\alpha t}\int_0^t e^{\alpha s}\left(L(s) - \frac{1}{2}\sigma^2\right)ds\right|^m \leqslant e^{-m\alpha t}\left(\int_0^t e^{\alpha s}\left|L(s) - \frac{1}{2}\sigma^2\right|ds\right)^m \\ &\leqslant \left|K_\mu K_\varphi + \frac{1}{2}\sigma^2\right|^m e^{-m\alpha t}\left(\int_0^t e^{\alpha s}ds\right)^m = \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha}\right)^m (1 - e^{-\alpha t})^m. \end{aligned} \tag{A.53}$$

Further, by Burkholder-Davis-Gundy inequality, there exists a positive constant $C_{m/2}$, such that

$$\begin{aligned} &\sigma^m e^{-m\alpha t}\mathbb{E}\left[\left|\int_0^t e^{\alpha s}dB_s\right|^m\right] \leqslant \sigma^m e^{-m\alpha t}\mathbb{E}\left[\left(\sup_{0\leqslant s\leqslant t}\left|\int_0^s e^{\alpha s}dB_s\right|\right)^m\right] \\ &\leqslant C_{m/2}\sigma^m e^{-m\alpha t}\mathbb{E}\left[\left(\int_0^t e^{2\alpha s}ds\right)^{m/2}\right] = C_{m/2}\sigma^m e^{-m\alpha t}\left(\frac{1}{2\alpha}(e^{2\alpha t} - 1)\right)^{m/2} \\ &= C_{m/2}\sigma^m\left(\frac{1}{2\alpha}(1 - e^{-2\alpha t})\right)^{m/2}, \end{aligned} \tag{A.54}$$

which implies that

$$\sup_{t\geq 0}\mathbb{E}\left[|\ln X(t)|^m\right] \leqslant 3^{m-1}\left(\mathbb{E}\left[|\ln X_0|^m\right] + \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha}\right)^m + C_{m/2}\sigma^m\left(\frac{1}{2\alpha}\right)^{m/2}\right) < \infty. \tag{A.55}$$

(2). From Proposition 2.2.1,

$$\mathbb{E}[X(t)] = \mathbb{E}\left[\exp\left\{e^{-\alpha t}\ln X_0 + e^{-\alpha t}\int_0^t e^{\alpha s}(L(s) - \frac{1}{2}\sigma^2)ds + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right\}\right]$$

By Assumption 2.2 and Proposition A.12,

$$\begin{aligned}\mathbb{E}[X(t)] &= \mathbb{E}\left[\exp\left\{e^{-\alpha t}\ln X_0\right\}\right]\exp\left\{e^{-\alpha t}\int_0^t e^{\alpha s}(L(s)-\frac{1}{2}\sigma^2)ds\right\}\mathbb{E}\left[\exp\left\{\sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right\}\right]\\
&= \mathbb{E}\left[\exp\left\{e^{-\alpha t}\ln X_0\right\}\right]\exp\left\{e^{-\alpha t}\int_0^t e^{\alpha s}(L(s)-\frac{1}{2}\sigma^2)ds\right\}\exp\left\{\frac{\sigma^2}{2}\frac{1}{2\alpha}(1-e^{-2\alpha t})\right\}\\
&\leqslant \mathbb{E}\left[X_0^{e^{-\alpha t}}\right]\exp\left\{e^{-\alpha t}\int_0^t e^{\alpha s}(L(s)-\frac{1}{2}\sigma^2)ds\right\}\exp\left\{\frac{\sigma^2}{2}\frac{1}{2\alpha}(1-e^{-2\alpha t})\right\}\end{aligned}$$

Since $0 < e^{-\alpha t} < 1$, by Jensen's inequality $\mathbb{E}\left[X_0^{e^{-\alpha t}}\right] \leqslant \{\mathbb{E}[X_0]\}^{e^{-\alpha t}}$, and this implies that

$$\begin{cases}\{\mathbb{E}[X_0]\}^{e^{-\alpha t}} \leqslant \mathbb{E}[X_0], & \text{if } \mathbb{E}[X_0] > 1\\ \{\mathbb{E}[X_0]\}^{e^{-\alpha t}} \leqslant 1, & \text{if } \mathbb{E}[X_0] \leqslant 1\end{cases}$$

then, $\{\mathbb{E}[X_0]\}^{e^{-\alpha t}} \leqslant \max\{\mathbb{E}[X_0], 1\}$

$$\mathbb{E}[X(t)] \leqslant \max\{\mathbb{E}[X_0], 1\}\exp\left\{e^{-\alpha t}\int_0^t e^{\alpha s}\left(L(s)-\frac{1}{2}\sigma^2\right)ds\right\}\exp\left\{\frac{\sigma^2}{2}\frac{1}{2\alpha}(1-e^{-2\alpha t})\right\}.$$

From (A.53) with $m = 1$, $e^{-\alpha t}\int_0^t e^{\alpha s}\left|L(s)-\frac{1}{2}\sigma^2\right|ds \leqslant \frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha}(1-e^{-\alpha t})$. Since $y = e^x$ is an increasing function of x for $x \geq 0$,

$$\begin{aligned}0 < \sup_{t\geq 0}\mathbb{E}[X(t)] &\leqslant \max\{\mathbb{E}[X_0], 1\}\sup_{t\geq 0}\exp\left\{\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha}(1-e^{-\alpha t}) + \frac{\sigma^2}{2}\frac{1}{2\alpha}(1-e^{-2\alpha t})\right\}\\
&\leqslant \max\{\mathbb{E}[X_0], 1\}\exp\left\{\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha}\sup_{t\geq 0}(1-e^{-\alpha t}) + \frac{\sigma^2}{2}\frac{1}{2\alpha}\sup_{t\geq 0}(1-e^{-2\alpha t})\right\}\\
&= \max\{\mathbb{E}[X_0], 1\}\exp\left\{\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha} + \frac{\sigma^2}{2}\frac{1}{\alpha}\right\} < \infty\end{aligned}$$

(3). From Proposition 2.2.1,

$$X(t)^m = \exp\left\{me^{-\alpha t}\ln X_0 + me^{-\alpha t}\int_0^t e^{\alpha s}\left(L(s)-\frac{1}{2}\sigma^2\right)ds + m\sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right\},$$

then, $\mathbb{E}[X(t)^m] = \mathbb{E}\left[\exp\left\{me^{-\alpha t}\ln X_0 + me^{-\alpha t}\int_0^t e^{\alpha s}\left(L(s)-\frac{1}{2}\sigma^2\right)ds + m\sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right\}\right]$.

By the independence between X_0 and $\{B_t, t \geq 0\}$ in Assumption 2.2,

$$\mathbb{E}[X(t)^m] = \mathbb{E}\left[\exp\left\{me^{-\alpha t}\ln X_0\right\}\right]\exp\left\{me^{-\alpha t}\int_0^t e^{\alpha s}\left(L(s)-\frac{1}{2}\sigma^2\right)ds\right\}\times$$

$$\mathbb{E}\left[\exp\left\{m\sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right\}\right].$$

By Jensen's inequality, and the fact $0 < e^{-\alpha t} < 1$,

$$\mathbb{E}\left[\exp\left\{me^{-\alpha t}\ln X_0\right\}\right] = \mathbb{E}\left[X_0^{me^{-\alpha t}}\right] \leqslant \{\mathbb{E}\left[X_0^m\right]\}^{e^{-\alpha t}} \leqslant \max\{\mathbb{E}\left[X_0^m\right], 1\}.$$

By Proposition A.12,

$$\begin{aligned}\mathbb{E}\left[\exp\left\{m\sigma e^{-\alpha t}\int_0^t e^{\alpha s}dB_s\right\}\right] &= \exp\left\{\tfrac{m^2\sigma^2 e^{-2\alpha t}}{2}\int_0^t e^{2\alpha s}ds\right\}\\ &= \exp\left\{\tfrac{m^2\sigma^2}{2}\tfrac{1}{2\alpha}(1-e^{-2\alpha t})\right\},\end{aligned}$$

which implies that

$$\begin{aligned}\mathbb{E}[X^m(t)] \leqslant \max\{\mathbb{E}\left[X_0^m\right], 1\}\exp\left\{me^{-\alpha t}\int_0^t e^{\alpha s}\left(L(s)-\frac{1}{2}\sigma^2\right)ds\right\}\times\\ \exp\left\{\frac{m^2\sigma^2}{2}\frac{1}{2\alpha}(1-e^{-2\alpha t})\right\}\\ < \max\{\mathbb{E}\left[X_0^m\right], 1\}\exp\left\{me^{-\alpha t}\int_0^t e^{\alpha s}|L(s)|ds\right\}\exp\left\{\frac{m^2\sigma^2}{2}\frac{1}{2\alpha}(1-e^{-2\alpha t})\right\}.\end{aligned}$$

Then, by the increasing property of the exponential function

$$\begin{aligned}0 < \sup_{t\geq 0}\mathbb{E}[X^m(t)] \leqslant \max\left\{\mathbb{E}\left[X_0^m\right], 1\right\}\exp\left\{m\sup_{t\geq 0}e^{-\alpha t}\int_0^t e^{\alpha s}|L(s)|ds\right\}\times\\ \exp\left\{\frac{m^2\sigma^2}{2}\frac{1}{2\alpha}\sup_{t\geq 0}(1-e^{-2\alpha t})\right\}\\ = \max\left\{\mathbb{E}\left[X_0^m\right], 1\right\}\exp\left\{m\frac{K_\mu K_\varphi}{\alpha}\right\}\exp\left\{\frac{m^2\sigma^2}{2}\frac{1}{2\alpha}\right\}.\end{aligned}$$

So, the conclusion that

$$\sup_{t\geq 0}\mathbb{E}[X^m(t)] \leqslant \max\left\{\mathbb{E}\left[X_0^m\right], 1\right\}\exp\left\{m\frac{K_\mu K_\varphi}{\alpha}+\frac{m^2\sigma^2}{2}\frac{1}{2\alpha}\right\} < \infty. \tag{A.56}$$

(4). Since

$$\left(\frac{S(\theta,t,X(t))}{\sigma X(t)}\right)^2 = ((\varphi(t), -\ln X(t))\theta)^\top(\varphi(t), -\ln X(t))\theta,$$

$$\begin{aligned}\left|\frac{S(\theta,t,X(t))}{\sigma X(t)}\right|^m = \left(|(\varphi(t), -\ln X(t))\theta)|^2\right)^{m/2}\\ \leqslant \left(\|(\varphi(t), -\ln X(t))\|^2\|\theta\|^2\right)^{m/2} = \left(\|\varphi(t)\|^2 + |\ln X(t)|^2\right)^{m/2}\|\theta\|^m.\end{aligned}$$

Since $m \geq 2$, $\left(\|\varphi(t)\|^2 + |\ln X(t)|^2\right)^{m/2} \leqslant 2^{m/2-1}\left(\|\varphi(t)\|^m + |\ln X(t)|^m\right)$, and

$$\mathbb{E}\left[\left|\frac{S(\theta,t,X(t))}{\sigma X(t)}\right|^m\right] \leqslant \|\theta\|^m 2^{m/2-1}\left(\|\varphi(t)\|^m + \mathbb{E}\left[|\ln X(t)|^m\right]\right).$$

From Proposition 2.2.2, $\sup_{t\geq 0}\mathbb{E}[|\ln X(t)|^m] < \infty$. This implies that

$$\mathbb{E}\left[\left|\frac{S(\theta,t,X(t))}{\sigma X(t)}\right|^m\right] < \|\theta\|^m 2^{m/2-1}\left(\|\varphi(t)\|^m + \sup_{t\geq 0}\mathbb{E}[|\ln X(t)|^m]\right) < \infty.$$

Then, $\sup_{t\geqslant 0}\mathbb{E}\left[\int_0^T \left|\frac{S(\theta,t,X(t))}{\sigma X(t)}\right|^m dt\right] < \infty$.

(5). $\mathrm{P}\left(\int_0^T \left|\frac{S(\theta,t,X(t))}{\sigma X(t)}\right|^m dt < \infty\right) = 1$ follows from Part **(4)** directly. This completes the proof. □

A.2.1 Other important results

Theorem A.1. *Theorem 3.5.8 [Stout, 1974] Let* $\{X_i, i \geqslant 1\}$ *be stationary and ergodic and let* ϕ *be a measurable function* $\phi: \mathbb{R}^\infty \mapsto \mathbb{R}^1$. *Let* $Y_i = \phi(X_i, X_{i+1}, \cdots)$ *and define* $\{Y_i, i \geqslant 1\}$. *Then,* $\{Y_i, i \geqslant 1\}$ *is stationary and ergodic.*

Proposition A.13. *Let* f *be bounded and Riemann-integrable on compact set* $[a,b]$. *Define a function* $C : [0,\infty) \to [0,\infty)$ *such that* (i) C *is non decreasing and* (ii) $\inf_{\delta>0} C(\delta) = C(0) = 0$. *Then, for all* x *in the dense subinterval of* $[a,b]$ *with continuity of* f, *there exists* $\delta > 0$, *such that* $|f(x\pm\delta) - f(x)| < C(\delta)$.

Proof. Suppose there exists some x_0 in the dense subinterval of $[a,b]$, $\exists\, \epsilon_0 > 0$, for all $\delta > 0$,

$$|f(x_0 \pm \delta) - f(x_0)| \geqslant \epsilon_0 > C(\delta). \tag{A.57}$$

For the constructed nested interval $I_n = [a_n, b_n]$, which contain x_0 as its unique limit point, $w^f_{I_n} < \frac{1}{2^{n-1}}$. This is a contradiction with (A.57) for the case of $x_0 \pm \delta \in I_n$. This completes the proof. □

Lemma A.4. *Let* $\{a_n\}_{n=0}^{\infty}$ *be a sequence of random variables and almost surely strictly positive, such that* $a_n \xrightarrow[n\to\infty]{a.s.} a$*, where* a *is a strictly positive non-random real number. Then,* a_n *is bounded away from 0 with probability 1.*

Proof. From $a_n \xrightarrow[n\to\infty]{a.s.} a$,, $\mathrm{P}\left(\left\{\omega : \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\sup_{2^k\leqslant u\leqslant 2^{k+1}} |a_u(\omega) - a| < a/2\right\}\right\}\right) = 1$. Thus, $\lim_{n\to\infty} \mathrm{P}\left(\left\{\omega : \bigcap_{k=n}^{\infty}\left\{\sup_{2^k\leqslant u\leqslant 2^{k+1}} a_u(\omega) > a/2\right\}\right\}\right) = 1$. This completes the proof. □

Lemma A.5. *Let* $\{A_t, t \in [0, T]\}$ *be a collection of* $p \times p$ *symmetric positive definite matrix, with random entries, and suppose that* $A_T \xrightarrow[T\to\infty]{a.s.} A$*, where* A *is a non-random symmetric positive definite matrix. Then the eigenvalues of the matrix* A_T *are bounded away from 0 and further, the smallest eigenvalue of* A_n *converges to the smallest eigenvalue of* A *with probability 1.*

Proof. Since $A_T \xrightarrow[T\to\infty]{a.s.} A$, which implies that $\|A_T\| \xrightarrow[T\to\infty]{a.s.} \|A\|$. Let $\lambda_{\max}(A_T)$ be the largest eigenvalue of the matrix A_T and $\lambda_{\max}(A)$ be the largest eigenvalue of the matrix A. that $\lambda_{\max}(A_T) \xrightarrow[T\to\infty]{a.s.} \lambda_{\max}(A)$ and since A_T, A are positive definite matrices, which implies that $A_T^{-1} \xrightarrow[T\to\infty]{a.s.} A^{-1}$. Then, $\lambda_{\max}(A_T^{-1}) \xrightarrow[T\to\infty]{a.s.} \lambda_{\max}(A^{-1})$. This is $\frac{1}{\lambda_{\max}(A_T^{-1})} \xrightarrow[T\to\infty]{a.s.} \frac{1}{\lambda_{\max}(A^{-1})}$, which is $\lambda_{\min}(A_T) \xrightarrow[T\to\infty]{a.s.} \lambda_{\min}(A)$. In addition, A is a symmetric positive definite matrix, which implies that $\lambda_{\min}(A)$ is strictly positive. By Lemma A.4, $\lambda_{\min}(A_T) \xrightarrow[n\to\infty]{a.s.} \lambda_{\min}(A)$ implies that $\lambda_{\min}(A_T)$ is almost surely bounded away from 0. Further, since $\lambda_{\min}(A_T)$ is the smallest eigenvalue, that all eigenvalues of A_T are bounded away from 0 with probability 1. This completes the proof. □

A.2.2 Properties of auxiliary process

The main challenge of the GEOU process consists in the fact that the process $\{X(t), t \geqslant 0\}$ is not stationary. In this subsection, to overcome this difficulty, an auxiliary process is introduced. The auxiliary process is strictly stationary and ergodic. Let $\tilde{B}_s =$

$B_s\mathbb{I}_{\mathbb{R}+}(s) + \bar{B}_{-s}\mathbb{I}_{\mathbb{R}-}(s)$ be a bilateral Brownian motion, where $\{B_s\}_{s\geqslant 0}$ and $\{\bar{B}_{-s}\}_{s\geqslant 0}$ are two independent Brownian motions. It is convenient to introduce an auxiliary process $\{\tilde{X}(t), t \geqslant 0\}$, which is given as

$$\ln \tilde{X}(t) = \tilde{r}(t) + \tilde{\tau}(t) \tag{A.58}$$

where $\tilde{r}(t) = e^{-\alpha t}\int_{-\infty}^{t} e^{\alpha s}\left(L(s) - \frac{1}{2}\sigma^2\right)ds$, $\tilde{\tau}(t) = \sigma e^{-\alpha t}\int_{-\infty}^{t} e^{\alpha s}d\bar{B}_s$.

Proposition A.14. *If Assumption 2.1-2.3 hold, then, for $k \in \mathbb{N}_+$,*

(1) $\mathbb{E}[\ln \tilde{X}(t+k-1)] = \tilde{r}(t)$*, for* $t \in [0,1]$*;*

(2) $\mathrm{Cov}(\ln \tilde{X}(t), \ln \tilde{X}(t+k)) = e^{-\alpha k}\frac{\sigma^2}{2\alpha}$*, for* $t \in [0,1]$*.*

Proof. **(1)** For $k = 1, 2, \cdots$, $\mathbb{E}[\ln \tilde{X}(t+k)] = \mathbb{E}[\tilde{r}(t+k)] + \mathbb{E}[\tilde{\tau}(t+k)] = \tilde{r}(t+k) + \mathbb{E}[\tilde{\tau}(t+k)]$ and $\tilde{r}(t+k) = e^{-\alpha(t+k)}\sum_{i=1}^{p}\theta_i\int_{-\infty}^{t+k} e^{\alpha s}\varphi_i(s)ds - \frac{1}{2}\sigma^2 e^{-\alpha(t+k)}\int_{-\infty}^{t+k} e^{\alpha s}ds$. Let $u = s - k$, by using Assumption 3, $\tilde{r}(t+k-1) = \tilde{r}(t)$. Further,

$$\mathbb{E}\left[\int_{-\infty}^{0} e^{\alpha s}d\bar{B}_{-s}\right] = \mathbb{E}\left[\int_{0}^{\infty} e^{-\alpha u}d\bar{B}_u\right] = \mathbb{E}\left[\lim_{U\to\infty}\int_{0}^{U} e^{-\alpha u}d\bar{B}_u\right] = \mathbb{E}\left[\lim_{U\to\infty} G_U\right]$$

where $G_U = \int_0^U e^{-\alpha s}d\bar{B}_s$. By Itô's isometry,

$$\mathbb{E}[G_U^2] = \mathbb{E}\left[\left(\int_0^U e^{-\alpha s}d\bar{B}_s\right)^2\right] = \mathbb{E}\left[\int_0^U e^{-2\alpha s}ds\right] \leqslant \frac{1}{2\alpha}(1 - e^{-2\alpha U})$$

which is bounded in U on $[0,\infty)$. Thus, by L^2 Bounded Martingale Convergence Theorem,

$$G_U \xrightarrow[U\to\infty]{L^2} G_\infty = \int_0^\infty e^{-\alpha s}d\bar{B}_s \tag{A.59}$$

which implies that $G_U \xrightarrow[U\to\infty]{L^1} G_\infty$. Then, $\lim_{U\to\infty}\mathbb{E}[G_U] = \mathbb{E}\left[\lim_{U\to\infty} G_U\right] = 0$, and then, $\mathbb{E}[\tilde{\tau}(t+k)] = 0$, for all $t \in [0,1]$, $k = 1, 2, \cdots$. Therefore, $\mathbb{E}[\ln \tilde{X}(t+k)] = \tilde{r}(t+k) + \mathbb{E}[\tilde{\tau}(t+k)] = \tilde{r}(t)$.

(2) Further, for the covariance,

$$\mathrm{Cov}(\ln \tilde{X}(t), \ln \tilde{X}(t+k)) = \mathrm{Cov}(\tilde{\tau}(t), \tilde{\tau}(t+k)) = \mathbb{E}[\tilde{\tau}(t)\tilde{\tau}(t+k)]$$

$$= \sigma^2 e^{-2\alpha t} e^{-\alpha k} \mathbb{E}\left[\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right)^2 + \left(\int_0^t e^{\alpha s} dB_s\right)^2 + 2\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right)\left(\int_0^t e^{\alpha s} dB_s\right)\right].$$

Since $\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right)$ and $\left(\int_0^t e^{\alpha s} dB_s\right)$ are independent,

$$\mathbb{E}\left[\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right)\left(\int_0^t e^{\alpha s} dB_s\right)\right] = \mathbb{E}\left[\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right)\right]\mathbb{E}\left[\left(\int_0^t e^{\alpha s} dB_s\right)\right] = 0.$$

Then,

$$\text{Cov}(\ln \tilde{X}(t), \ln \tilde{X}(t+k)) = \sigma^2 e^{-2\alpha t} e^{-\alpha k} \mathbb{E}\left[\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right)^2\right]$$

$$+\sigma^2 e^{-2\alpha t} e^{-\alpha k} \mathbb{E}\left[\left(\int_0^t e^{\alpha s} dB_s\right)^2\right],$$

and then, from (A.59), $\text{Cov}(\ln \tilde{X}(t), \ln \tilde{X}(t+k)) = e^{-\alpha k} \frac{\sigma^2}{2\alpha}$.

This completes the proof. □

Proposition A.14 implies that the process $\{\ln \tilde{X}(t+k-1), k \in \mathbb{N}_+\}$ is a wide sense stationary process for each $t \in [0,1]$. The following proposition proves that the sequence of random variables $\left\{\ln \tilde{X}(t+k-1), t \in [0,1]\right\}_{k=1}^{\infty}$ is stationary in strict sense and ergodic.

Proposition A.15. *For $k \in \mathbb{N}_+$, the sequence of random variables $\{\ln \tilde{X}(t+k-1), t \in [0,1]\}$ is stationary and ergodic.*

Proof. Let $\tilde{\tau}(t) = \sigma e^{-\alpha t}\left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s} + \int_0^t e^{\alpha s} dB_s\right)$. First, to derive the distribution of $\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}$, let $u = -s$, $\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s} = -\int_0^{\infty} e^{-\alpha u} d\bar{B}_u$, and $\mathscr{Y}_T = \int_0^T e^{-\alpha u} d\bar{B}_u$. From Proposition A.12, the *Moment Generating Function* of $\mathscr{Y}_T$ is

$$\mathscr{M}_{\mathscr{Y}_T}(u) = \exp\left\{\frac{1}{2}\frac{1}{2\alpha}(1-e^{-2\alpha T})u^2\right\}, \quad u \in \mathbb{R},$$

which implies that

$$\lim_{T\to+\infty} \mathscr{M}_{\mathscr{Y}_T}(u) = \lim_{T\to\infty} \exp\left\{\frac{1}{2}\frac{1}{2\alpha}(1-e^{-2\alpha T})u^2\right\} = \exp\left\{\frac{1}{2}\frac{1}{2\alpha}u^2\right\}, \quad u \in \mathbb{R}.$$

Then,

$$\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s} \sim \mathcal{N}\left(0, \frac{1}{2\alpha}\right). \tag{A.60}$$

Further, by Proposition A.12, that $\int_0^t e^{\alpha s} dB_s \sim \mathcal{N}\left(0, \frac{1}{2\alpha}(e^{2\alpha t} - 1)\right)$. By the independent between $\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}$ and $\int_0^t e^{\alpha s} dB_s$,

$$\tilde{\tau}(t) = \sigma e^{-\alpha t} \left(\int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s} + \int_0^t e^{\alpha s} dB_s \right) \sim \mathcal{N}\left(0, \frac{\sigma^2}{2\alpha}\right),$$

which implies that for $k = 1, 2, \cdots$, $\ln \tilde{X}(t + k - 1) \sim \mathcal{N}\left(\tilde{h}(t), \frac{\sigma^2}{2\alpha}\right)$, i.e. the process $\left\{\ln \tilde{X}(t + k - 1), 0 \le t \le 1\right\}_{k=0}^{\infty}$ is stationary in strict sense. From the second part of Proposition A.14, for $t \in [0, 1]$, $k \in \{1, 2, 3, \cdots\}$, the covariance of $\ln \tilde{X}(t), \ln \tilde{X}(t + k)$ is

$$R_k = \mathrm{Cov}(\ln \tilde{X}(t), \ln \tilde{X}(t + k)) = e^{-\alpha k} \frac{\sigma^2}{2\alpha},$$

which implies that $R_k \to 0$ as $k \to \infty$. From Example 3.5.2 in Stout [1974], it concludes that the process $\left\{\ln \tilde{X}(t + k - 1), 0 \le t \le 1\right\}_{k=0}^{\infty}$ is stationary and ergodic. □

To study the long term behavior of the solution of the GEOU by using the stationary and ergodic property of the auxiliary process $\{\ln \tilde{X}(t + k), 0 \leqslant t \leqslant 1\}_{k=0}^{\infty}$, below, the result gives the relationship between $\ln X(t)$ and $\ln \tilde{X}(t)$ when t is large. The following proposition shows that $\ln \tilde{X}(t) - \ln X(t)$ converges in L^m and almost surely to 0 as t tends to infinity.

Proposition A.16. *If Assumption 2.1-2.3 hold, then, (1)* $|\ln \tilde{X}(t) - \ln X(t)| \xrightarrow[T\to\infty]{a.s.\text{ and } L^m} 0$*; (2)* $(\ln X(t))^2 - (\ln \tilde{X}(t))^2 \xrightarrow[t\to\infty]{a.s.\text{ and } L^{m/2}} 0$.

Proof. **(1)**. First, it is obvious that

$$|\ln \tilde{X}(t) - \ln X(t)| \leqslant |e^{-\alpha t} \ln X_0| + e^{-\alpha t} \left| \int_{-\infty}^0 e^{\alpha s} \left(L(s) - \frac{1}{2}\sigma^2 \right) ds \right| + \sigma e^{-\alpha t} \left| \int_{-\infty}^0 e^{\alpha s} d\tilde{B}_s \right|.$$

(a) By Assumption 2.2, $X_0 > 0$ a.s. and $\mathbb{E}(|X_0|^m) < \infty$,, $|e^{-\alpha t} \ln X_0| = e^{-\alpha t} |\ln X_0| \xrightarrow[t\to\infty]{a.s.} 0$.

(b) $\left| \int_{-\infty}^0 e^{\alpha s} \left(L(s) - \frac{1}{2}\sigma^2 \right) ds \right| \leqslant \frac{|K_\mu K_\varphi|^2}{\alpha}$ implies that $e^{-\alpha t} \left| \int_{-\infty}^0 e^{\alpha s} \left(L(s) - \frac{1}{2}\sigma^2 \right) ds \right| \xrightarrow[t\to\infty]{} 0$.

(c) $\int_{-\infty}^{0} e^{\alpha s} d\tilde{B}_{-s}$ is a random variable which is independent with t, and $\mathbb{E}\left[\left|\int_{-\infty}^{0} e^{\alpha s} d\tilde{B}_{-s}\right|^2\right] < \infty$, then $\sigma e^{-\alpha t}\left|\int_{-\infty}^{0} e^{\alpha s} d\tilde{B}_s\right| \xrightarrow[t\to\infty]{a.s.} 0$.

This completes the proof of convergence almost surely. Second, the convergence in L^m is proven.

$$\mathbb{E}\left[|\ln \tilde{X}(t) - \ln X(t)|^m\right] \leqslant 3^{m-1}\left(e^{-m\alpha t}\mathbb{E}\left[|\ln X_0|^m\right] + e^{-m\alpha t}\left|\int_{-\infty}^{0} e^{\alpha s}\left(L(s) - \frac{1}{2}\sigma^2\right) ds\right|^m + \sigma^2 e^{-m\alpha t}\mathbb{E}\left[\left|\int_{-\infty}^{0} e^{\alpha s} d\tilde{B}_s\right|^m\right]\right)$$

(A) From Assumption 2.2, $3^{m-1}e^{-m\alpha t}\mathbb{E}\left[|\ln X_0|^m\right] \xrightarrow[t\to\infty]{} 0$.

(B) Since $L(t)$ is bounded,$3^{m-1}e^{-m\alpha t}\left|\int_{-\infty}^{0} e^{\alpha s}\left(L(s) - \frac{1}{2}\sigma^2\right) ds\right|^m \xrightarrow[t\to\infty]{} 0$.

(C) By L^m-bounded martingale convergence theorem, $|G_U| \xrightarrow[U\to\infty]{L^m} |G_\infty| = \left|\int_0^{\infty} e^{-\alpha s} d\bar{B}_s\right|$, and then,

$$3^{m-1}\sigma^2 e^{-m\alpha t}\mathbb{E}\left[\left|\int_{-\infty}^{0} e^{\alpha s} d\tilde{B}_s\right|^m\right] \leqslant 3^{m-1}\sigma^2 e^{-m\alpha t} C_{m/2}\left(\frac{1}{2\alpha}\right)^{m/2} \xrightarrow[t\to\infty]{} 0.$$

This completes the proof of convergence in L^m.

(2) For the second assertion,

$$\begin{aligned}\left|(\ln X(t))^2 - (\ln \tilde{X}(t))^2\right|^{m/2} &\leqslant 2^{m/2-1}\Big(|\ln X(t)|^{m/2}\left|\ln X(t) - \ln \tilde{X}(t)\right|^{m/2} \\ &\quad + \left|\ln \tilde{X}(t)\right|^{m/2}\left|\ln X(t) - \ln \tilde{X}(t)\right|^{m/2}\Big),\end{aligned} \tag{A.61}$$

and by Cauchy-Schwartz inequality,

$$\mathbb{E}\left[|\ln X(t)|^{m/2}\left|\ln X(t) - \ln \tilde{X}(t)\right|^{m/2}\right] \leqslant \sqrt{\mathbb{E}\left[|\ln X(t)|^m\right]\mathbb{E}\left[\left|\ln X(t) - \ln \tilde{X}(t)\right|^m\right]}.$$

By (A.55), $\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right] < \infty$, and by Proposition A.16, $\mathbb{E}\left[|\ln \tilde{X}(t) - \ln X(t)|^m\right] \xrightarrow[t\to\infty]{} 0$, then

$$\mathbb{E}\left[|\ln X(t)|^{m/2}\left|\ln X(t) - \ln \tilde{X}(t)\right|^{m/2}\right] \leqslant \sqrt{\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\sqrt{\mathbb{E}\left[\left|\ln X(t) - \ln \tilde{X}(t)\right|^m\right]} \xrightarrow[t\to\infty]{} 0. \tag{A.62}$$

Further, by (A.58), $\left(\ln \tilde{X}(t)\right)^m = (\tilde{r}(t) + \tilde{\tau}(t))^m \leqslant 2^{m-1}\left(\tilde{r}^m(t) + \tilde{\tau}^m(t)\right)$, and $\mathbb{E}[|\tilde{\tau}(t)|^m] \leqslant C_{m/2}\left(\frac{1}{2\alpha}\right)^{m/2}\left(e^{-\alpha m t} + \left(1 - e^{-2\alpha t}\right)^{m/2}\right) \leqslant 2C_{m/2}\left(\frac{1}{2\alpha}\right)^{m/2}$. Then,

$$\sup_{t\geqslant 0}\mathbb{E}[|\ln \tilde{X}(t)|^m] \leqslant 2^{m-1}\left(\left(\frac{K_\varphi K_\mu}{\alpha}\right)^m + 2C_{m/2}\left(\frac{1}{2\alpha}\right)^{m/2}\right) < \infty. \tag{A.63}$$

Apply Proposition A.16 again, and (A.63),

$$\mathbb{E}\left[\left|\ln \tilde{X}(t)\right|^{m/2}\left|\ln X(t) - \ln \tilde{X}(t)\right|^{m/2}\right] \leqslant \sqrt{\sup_{t\geqslant 0}\mathbb{E}\left[\left|\ln \tilde{X}(t)\right|^m\right]}\sqrt{\mathbb{E}\left[\left|\ln X(t) - \ln \tilde{X}(t)\right|^m\right]} \xrightarrow[t\to\infty]{} 0. \tag{A.64}$$

(A.61), (A.62) and (A.64) complete the proof of assertion (2) with convergence in L^m.

Second, by triangle inequality,

$$\left|(\ln X(t))^2 - \ln(\tilde{X}(t))^2\right| \leqslant |\ln X(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| + \left|\ln \tilde{X}(t)\right|\left|\ln X(t) - \ln \tilde{X}(t)\right|. \tag{A.65}$$

Since $\left|\ln \tilde{X}(t)\right|\left|\ln X(t) - \ln \tilde{X}(t)\right| \leqslant |\tilde{r}(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| + |\tilde{\tau}(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right|$, and $\tilde{r}(t)$ is bounded,

$$|\tilde{r}(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| \xrightarrow[t\to\infty]{a.s.} 0. \tag{A.66}$$

Further, since $\int_0^t e^{\alpha s} d\tilde{B}_s$ is a martingale, from Doob's maximal inequality for sub-martingales. For any $\varepsilon > 0$,

$$\begin{aligned} P\left(\sup_{k\leqslant t\leqslant k+1} e^{-2\alpha t}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right| > \varepsilon\right) &\leqslant P\left(\sup_{k\leqslant t\leqslant k+1}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right| > \varepsilon e^{2\alpha k}\right) \\ &\leqslant \frac{\mathbb{E}\left[\sup_{k\leqslant t\leqslant k+1}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right|^2\right]}{\varepsilon^2 e^{4\alpha k}}. \end{aligned}$$

By Burkholder-Davis-Gundy inequality,

$$P\left(\sup_{k\leqslant t\leqslant k+1} e^{-2\alpha t}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2 e^{4\alpha k}} 4\mathbb{E}\left[\left|\int_0^{k+1} e^{2\alpha s} ds\right|\right].$$

Further, $\mathbb{E}\left[\left|\int_0^{k+1} e^{\alpha s} d\tilde{B}_s\right|^2\right] = \frac{1}{2\alpha}(e^{2\alpha(k+1)} - 1) \leqslant \frac{1}{2\alpha} e^{2\alpha(k+1)}$, which implies that

$$P\left(\sup_{k\leqslant t\leqslant k+1} e^{-2\alpha t}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right| > \varepsilon\right) \leqslant \frac{1}{\varepsilon^2 e^{4\alpha k}}\frac{1}{2\alpha} 4e^{2\alpha(k+1)} = \frac{2e^{2\alpha}}{\alpha\varepsilon^2} e^{-2\alpha k}$$

and

$$\sum_{k=1}^{\infty} \mathrm{P}\left(\sup_{k\leqslant t\leqslant k+1} e^{-2\alpha t}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right| > \varepsilon\right) \leqslant \frac{2e^{2\alpha}}{\alpha\varepsilon^2}\sum_{k=1}^{\infty} e^{-2\alpha k} = \frac{2e^{2\alpha}}{\alpha\varepsilon^2}\frac{e^{-2\alpha}}{1-e^{-2\alpha}} < \infty.$$

Then, by Borel–Cantelli Lemma,

$$e^{-2\alpha t}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right| \xrightarrow[t\to\infty]{a.s.} 0. \tag{A.67}$$

Together with $e^{-2\alpha t}\left|\int_{-\infty}^0 e^{\alpha s} d\tilde{B}_s\right| \xrightarrow[t\to\infty]{a.s.} 0$, $e^{-2\alpha t}\left|\int_{-\infty}^t e^{\alpha s} d\tilde{B}_s\right| \xrightarrow[t\to\infty]{a.s.} 0$. This implies that

$$|\tilde{\tau}(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| \xrightarrow[t\to\infty]{a.s.} 0. \tag{A.68}$$

Further, by Proposition 2.2.1,

$$|\ln X(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| \leqslant e^{-\alpha t}\,|\ln X_0|\left|\ln X(t) - \ln \tilde{X}(t)\right|$$

$$+ |r(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| + |\tau(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right|.$$

Since $r(t)$ is bounded,

$$|r(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| \xrightarrow[t\to\infty]{a.s.} 0 \tag{A.69}$$

and from Proposition A.16,

$$e^{-\alpha t}\,|\ln X_0|\left|\ln X(t) - \ln \tilde{X}(t)\right| \xrightarrow[t\to\infty]{a.s.} 0. \tag{A.70}$$

Further, $\left|\ln X(t) - \ln \tilde{X}(t)\right| = e^{-\alpha t}\left|\ln X_0 + \int_{-\infty}^0 e^{\alpha s}(L(s) - \frac{1}{2}\sigma^2)ds + \int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right|$, implies

$$|\tau(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| = e^{-2\alpha t}\left|\int_0^t e^{\alpha s} d\tilde{B}_s\right|\left|\ln X_0 + \int_{-\infty}^0 e^{\alpha s}(L(s) - \frac{1}{2}\sigma^2)ds + \int_{-\infty}^0 e^{\alpha s} d\bar{B}_{-s}\right|.$$

From (A.67), $e^{-2\alpha t}\left|\int_0^t e^{\alpha s} dB_s\right| \xrightarrow[t\to\infty]{a.s.} 0$ and then

$$|\tau(t)|\left|\ln X(t) - \ln \tilde{X}(t)\right| \xrightarrow[t\to\infty]{a.s.} 0. \tag{A.71}$$

(A.65), (A.66), (A.68), (A.69), (A.70) and (A.71) imply that $(\ln X(t))^2 - (\ln \tilde{X}(t))^2 \xrightarrow[t\to\infty]{a.s.} 0$. This completes the proof. □

A.3 On parameter estimation and asymptotic results

A.3.1 Derivation of MLE

Proof of Proposition 2.3.1. Let $a = [a_1^\top, a_2]$ with a_1 a $p-$column vector, and a_2 a scalar. Then, $aQ_{[0,T]}a^\top = (a_1^\top, a_2)Q_{[0,T]}(a_1^\top, a_2)^\top = \int_0^T \left(a_1^\top \varphi^\top(t) - a_2 \ln X(t)\right)^2 dt \geqslant 0$, and the equality holds if and only if $\left(a_1^\top \varphi^\top(t) - a_2 \ln X(t)\right)^2 = 0$, almost everywhere for $t \in [0, T]$, which is $a_1^\top \varphi^\top(t) - a_2 \ln X(t) = 0$, almost everywhere for $t \in [0, T]$, i.e.

$$\mathrm{P}\left(\omega : a_1^\top \varphi^\top(t) - a_2 \ln X(t, \omega) = 0, \forall\, t \in [0, T]\right) = 1, \tag{A.72}$$

It is obvious that if $a_2 = 0$, by Assumption 2.3, $a_1 = \vec{0}_{p\times 1}$. First, $a_2 = 0$ is need to be proven. Note that, $\ln X(t)|X_0$ follows normal distribution, which implies that $\left(a_1^\top \varphi^\top(t) - a_2 \ln X(t)\right)|X_0$ also follows normal distribution, and

$$\mathrm{Var}\left(\left(a_1^\top \varphi^\top(t) - a_2 \ln X(t)\right)|X_0\right) = a_2^2 \frac{\sigma^2}{2\alpha}(1 - e^{-2\alpha t}) > 0,$$

for all $t \in [0, T]$. This shows that $\left(a_1^\top \varphi^\top(t) - a_2 \ln X(t)\right)|X_0$ is not a constant with probability 1, which means that for $a_2 \neq 0$,

$$\mathrm{P}\left(\omega : a_1^\top \varphi^\top(t) - a_2 \ln X(t,\omega) = 0, \forall\, t \in [0, T]\right) = 0.$$

This is a contradiction with (A.72), hence, $a_2 = 0$, which implies that $a_1^\top \varphi^\top(t) = 0$. By Assumption 2.3, provided $T \geqslant 1$, the base function $\varphi(t)$ is linearly independent, which means $a_1^\top \varphi^\top(t) = 0$ can imply that $a_1 = \vec{0}_{p\times 1}$ and $a = \vec{0}_{1\times(p+1)}$. This completes the proof. □

A.3.2 Joint asymptotic normality and related results

In this subsection, the joint asymptotic normality of the estimators is derived. The limiting distributions play an important role in deriving shrinkage estimators and their asymptotic relative efficiency. Further, the limiting distributions are also important in solving the hypothesis testing problem (2.2.3). To this end, the asymptotic behavior of the positive definite matrix $\frac{1}{T}Q_{[0,T]}$ and the column vector $\frac{1}{\sqrt{T}}W_{[0,T]}$ is studied. The preliminary propositions play crucial roles in deriving the limiting distributions of both UMLE and RMLE. The following proposition proves the result in a more general case.

Proposition A.17. *Let $\varphi^{\top}(t)$ be a p-column vector of real-valued functions, which are linearly independent and Riemann square integrable. Let $\tilde{r}(t)$ be square integrable real-valued function, $\Lambda = \int_0^1 \varphi^{\top}(t)\tilde{r}(t)dt$, $\beta \neq 0$ and let $\Pi = \begin{bmatrix} \int_0^1 \varphi^{\top}(t)\varphi(t)dt & -\Lambda \\ -\Lambda^{\top} & \int_0^1 \tilde{r}^2(t)dt + \beta^2 \end{bmatrix}$. Then, Π is a positive definite matrix.*

Proof. Let $a = (a_1^{\top}, a_2)^{\top}$, where a_1 is a p_0-column vector and a_2 is a scalar. Then, $a^{\top}\Pi\, a = a_1^{\top}\int_0^1 \varphi(t)\varphi^{\top}(t)dta_1 - 2a_1^{\top}\int_0^1 \widetilde{r}_j(t)\varphi(t)dta_2 + a_2\left(\int_0^1 \widetilde{r}_j^2(t)dt\right)a_2 + a_2^2\beta^2$. Then, $a^{\top}\Pi\, a = \int_0^1 \left(a_1^{\top}\varphi^{\top}(t) - a_2\widetilde{r}_j(t)\right)^2 dt + a_2^2\beta^2 \geqslant 0$. Further, $a^{\top}\Pi\, a = 0$ if and only if $\int_0^1 \left(a_1^{\top}\varphi^{\top}(t) - a_2\widetilde{r}_j(t)\right)^2 dt = 0, \quad a_2^2\beta^2 = 0$. Since $\beta^2 > 0$, this implies that $a_2 = 0$ and $\int_0^1 \left(a_1^{\top}\varphi^{\top}(t) - a_2\widetilde{r}_j(t)\right)^2 dt = 0$. Combining $a_2 = 0$ and $\int_0^1 \left(a_1^{\top}\varphi^{\top}(t) - a_2\widetilde{r}_j(t)\right)^2 dt = 0$, together with the fact that $\varphi^{\top}(t)$ is a p-column vector of linearly independent functions, $a_1 = \vec{0}_{p\times 1}$, which is $a^{\top}\Pi\, a = 0$ if and only if $a = \vec{0}_{(p+1)\times 1}$. This completes the proof. □

Proof of Proposition 2.3.4. This proposition will be proven in the following four steps:

(1) Step 1, by Assumption 2.3,

$$\frac{1}{T}\int_0^T \varphi^{\top}(t)\varphi(t)dt = \frac{1}{T}\int_0^{\lfloor T\rfloor} \varphi^{\top}(t)\varphi(t)dt + \frac{1}{T}\int_{\lfloor T\rfloor}^T \varphi^{\top}(t)\varphi(t)dt$$

$$= \frac{1}{T}\sum_{j=1}^{\lfloor T\rfloor}\int_{j-1}^{j} \varphi^{\top}(t)\varphi(t)dt + \frac{1}{T}\int_{\lfloor T\rfloor}^T \varphi^{\top}(t)\varphi(t)dt$$

By Assumption 2.3,

$$\frac{1}{T}\sum_{j=1}^{\lfloor T\rfloor}\int_{j-1}^{j} \varphi^{\top}(t)\varphi(t)dt = \frac{1}{T}\sum_{j=1}^{\lfloor T\rfloor}\int_0^1 \varphi^{\top}(u)\varphi(u)du = \frac{\lfloor T\rfloor}{T}I_p \xrightarrow[T\to\infty]{} I_p.$$

Further, by Jensen inequality, property of periodic function, let $u = t - \lfloor T \rfloor$,

$$\left\| \frac{1}{T} \int_{\lfloor T \rfloor}^{T} \varphi^{\top}(t)\varphi(t)dt \right\| = \frac{1}{T} \left\| \int_{0}^{T-\lfloor T \rfloor} \varphi^{\top}(u + \lfloor T \rfloor)\varphi(u + \lfloor T \rfloor)dt \right\|$$

$$\leq \frac{1}{T} \int_{0}^{1} \left\| \varphi^{\top}(u)\varphi(u) \right\| dt \xrightarrow[T\to\infty]{} 0,$$

which implies that $\frac{1}{T} \int_{\lfloor T \rfloor}^{T} \varphi^{\top}(t)\varphi(t)dt \xrightarrow[T\to\infty]{} 0$. Thus,

$$\frac{1}{T} \int_{0}^{T} \varphi^{\top}(t)\varphi(t)dt \xrightarrow[T\to\infty]{} I_p.$$

(2) Step 2, indeed, by Proposition A.16, $|\ln \tilde{X}(t) - \ln X(t)| \xrightarrow[T\to\infty]{a.s.\text{ and } L^m} 0$, together with $\|\varphi(t)\| \leq K_{\varphi}$,

$$\varphi^{\top}(t) \ln X(t) - \varphi^{\top}(t) \ln \tilde{X}(t) \xrightarrow[T\to\infty]{a.s.\text{ and } L^m} 0.$$

By the continuous version of Cesàro mean theorem,

$$\frac{1}{T} \int_{0}^{T} \varphi^{\top}(t) \ln X(t)dt - \frac{1}{T} \int_{0}^{T} \varphi^{\top}(t) \ln \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.\text{ and } L^m} 0.$$

Then,the following needs to be proven

$$\frac{1}{T} \int_{0}^{T} \varphi^{\top}(t) \ln \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.\text{ and } L^m} \int_{0}^{1} \varphi^{\top}(t)\tilde{r}(t)dt.$$

(i) In fact,

$$\frac{1}{T} \int_{0}^{T} \varphi^{\top}(t) \ln \tilde{X}(t)dt = \frac{1}{T} \int_{0}^{\lfloor T \rfloor} \varphi^{\top}(t) \ln \tilde{X}(t)dt + \frac{1}{T} \int_{\lfloor T \rfloor}^{T} \varphi^{\top}(t) \ln \tilde{X}(t)dt$$

$$= \frac{1}{T} \sum_{j=1}^{\lfloor T \rfloor} \int_{j-1}^{j} \varphi^{\top}(t) \ln \tilde{X}(t)dt + \frac{1}{T} \int_{\lfloor T \rfloor}^{T} \varphi^{\top}(t) \ln \tilde{X}(t)dt.$$

Let $Y_j = \int_{j-1}^{j} \varphi^{\top}(t) \ln \tilde{X}(t)dt$. By Proposition A.15, $\{\ln \tilde{X}(u + k - 1), 0 \leq u \leq 1\}_{k\in\mathbb{N}_+}$ is stationary and ergodic. Then, that $\{Y_j\}_{j\in\mathbb{N}_+}$ is stationary and ergodic. This leads to

$$\frac{1}{T} \sum_{j=1}^{\lfloor T \rfloor} \int_{j-1}^{j} \varphi^{\top}(t) \ln \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.\text{ and } L^{m/2}} \int_{0}^{1} \varphi^{\top}(u)\mathbb{E}\left[\ln \tilde{X}_u\right] du = \int_{0}^{1} \varphi(t)\tilde{r}(t)dt. \tag{A.73}$$

(ii) In addition, for $i = 1, 2, \cdots, p$, let $S_{\lfloor T\rfloor i} = \int_{\lfloor T\rfloor}^{T} \varphi_i(t) \ln \tilde{X}(t)dt$,

$$\mathbb{E}[S^2_{\lfloor T\rfloor i}] = \mathbb{E}\left[\left(\int_{\lfloor T\rfloor}^{T} \varphi_i(t) \ln \tilde{X}(t)dt\right)^2\right] \leq K_\varphi^2 \mathbb{E}\left[\left(\int_{\lfloor T\rfloor}^{T} \ln \tilde{X}(t)dt\right)^2\right]$$

$$= K_\varphi^2 \mathbb{E}\left[\left(\int_{\lfloor T\rfloor}^{T} \ln \tilde{X}(t) \int_{\lfloor T\rfloor}^{T} \ln \tilde{X}_u dudt\right)\right] = K_\varphi^2 \mathbb{E}\left[\left(\int_{\lfloor T\rfloor}^{T} \int_{\lfloor T\rfloor}^{T} \ln \tilde{X}(t) \ln \tilde{X}_u dudt\right)\right].$$

Then,

$$\mathbb{E}[S^2_{\lfloor T\rfloor i}] = K_\varphi^2 \left(\int_{\lfloor T\rfloor}^{T} \int_{\lfloor T\rfloor}^{T} \mathbb{E}\left[\ln \tilde{X}(t) \ln \tilde{X}_u\right] dudt\right).$$

By (A.63), for $m = 2$, $\mathbb{E}\left[\ln \tilde{X}(t) \ln \tilde{X}_u\right] \leq \sup_{t\geq 0} \mathbb{E}\left[\left|\ln \tilde{X}(t)\right|^2\right] < \infty$, then,

$$\mathbb{E}[S^2_{\lfloor T\rfloor i}] \leq K_\varphi^2 \sup_{t\geq 0} \mathbb{E}\left[|X(t)|^2\right] \left(\int_{\lfloor T\rfloor}^{T} \int_{\lfloor T\rfloor}^{T} dudt\right) = K_\varphi^2 \left(\sup_{t\geq 0} \mathbb{E}\left[|X(t)|^2\right]\right) (T - \lfloor T\rfloor)^2.$$

Since $T - \lfloor T\rfloor \leq 1$,

$$\sum_{\lfloor T\rfloor=1}^{\infty} \mathrm{P}\left(\frac{|S_{\lfloor T\rfloor i}|}{\lfloor T\rfloor} > \varepsilon\right) \leq \sum_{\lfloor T\rfloor=1}^{\infty} \frac{\mathbb{E}[|S_{\lfloor T\rfloor i}|^2]}{\lfloor T\rfloor^2 \varepsilon^2} \leq \sum_{\lfloor T\rfloor=1}^{\infty} \frac{K_\varphi^2 \left(\sup_{t\geq 0} \mathbb{E}\left[|X(t)|^2\right]\right)}{\lfloor T\rfloor^2 \varepsilon^2}$$

$$= K_\varphi^2 \left(\sup_{t\geq 0} \mathbb{E}\left[|X(t)|^2\right]\right) \sum_{\lfloor T\rfloor=1}^{\infty} \frac{1}{\lfloor T\rfloor^2 \varepsilon^2} < \infty,$$

by Borel-Cantelli Lemma, $\frac{|S_{\lfloor T\rfloor i}|}{\lfloor T\rfloor} \xrightarrow[T\to\infty]{a.s.} 0$, for $i = 1, 2, \cdots, p$, which implies that

$$\frac{1}{T} \int_{\lfloor T\rfloor}^{T} \varphi^\top(t) \ln \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.} 0. \tag{A.74}$$

Further,

$$\mathbb{E}\left[\left\|\frac{1}{T} \int_{\lfloor T\rfloor}^{T} \varphi^\top(t) \ln \tilde{X}(t)dt\right\|^m\right] = \mathbb{E}\left[\left\|\frac{1}{T} \int_{\lfloor T\rfloor}^{T} \varphi^\top(t) \ln \tilde{X}(t)dt\right\|^m\right]$$
$$\leq \mathbb{E}\left[\left(\frac{1}{T} \int_{\lfloor T\rfloor}^{T} \left\|\varphi^\top(t)\right\| |\ln \tilde{X}(t)|dt\right)^m\right] \leq \left(\frac{1}{T}\right)^m \left(pK_\varphi\right)^{m/2} \mathbb{E}\left[\left(\int_{\lfloor T\rfloor}^{T} |\ln \tilde{X}(t)|dt\right)^m\right]$$
$$\leq \left(\frac{1}{T}\right)^m \left(pK_\varphi\right)^{m/2} (T - \lfloor T\rfloor)^{m-1} \left(\int_{\lfloor T\rfloor}^{T} \mathbb{E}\left[|\ln \tilde{X}(t)|^m\right] dt\right)$$
$$\leq \left(\frac{1}{T}\right)^m \left(pK_\varphi\right)^{m/2} \left(\int_{\lfloor T\rfloor}^{T} \mathbb{E}\left[|\ln \tilde{X}(t)|^m\right] dt\right).$$

By (A.63), $\sup_{t\geq 0} \mathbb{E}\left[\left|\ln \tilde{X}(t)\right|^m\right] < \infty$, then,

$$\mathbb{E}\left[\left\|\frac{1}{T}\int_{\lfloor T\rfloor}^{T} \varphi^{\top}(t)\ln \tilde{X}(t)dt\right\|^m\right] \leq \left(\frac{1}{T}\right)^m \left(pK_{\varphi}\right)^{m/2} \left(\sup_{t\geq 0}\mathbb{E}\left[\left|\ln \tilde{X}(t)\right|^m\right]\right) \xrightarrow[T\to\infty]{} 0. \tag{A.75}$$

(A.73), (A.74) and (A.75) imply that

$$\frac{1}{T}\int_0^T \varphi(t)\ln X(t)dt \xrightarrow[T\to\infty]{a.s.\text{ and }L^m} \int_0^1 \varphi(t)\tilde{r}(t)dt.$$

(3) Step 3, from Proposition A.16, $(\ln X(t))^2 - (\ln \tilde{X}(t))^2 \xrightarrow[T\to\infty]{a.s.\text{ and }L^{m/2}} 0$. By the continuous version of Cesàro mean theorem,

$$\frac{1}{T}\int_0^T (\ln X(t))^2 dt - \frac{1}{T}\int_0^T (\ln \tilde{X}(t))^2 dt \xrightarrow[T\to\infty]{a.s.\text{ and }L^{m/2}} 0.$$

The following convergence needs to be proven

$$\frac{1}{T}\int_0^T (\ln \tilde{X}(t))^2 dt \xrightarrow[T\to\infty]{a.s.\text{ and }L^{m/2}} \int_0^1 (\tilde{r}(t))^2 dt + \frac{\sigma^2}{2\alpha}.$$

By the fact that

$$\frac{1}{T}\int_0^{\lfloor T\rfloor} (\ln \tilde{X}(t))^2 dt \leq \frac{1}{T}\int_0^{T} (\ln \tilde{X}(t))^2 dt \leq \frac{1}{T}\int_0^{\lfloor T\rfloor+1} (\ln \tilde{X}(t))^2 dt. \tag{A.76}$$

Then,

$$\text{LHS of } (A.76) = \frac{1}{T}\int_0^{\lfloor T\rfloor} (\ln \tilde{X}(t))^2 dt = \frac{1}{T}\sum_{j=1}^{\lfloor T\rfloor}\int_{j-1}^{j} (\ln \tilde{X}(t))^2 dt$$

By Proposition A.15, $\left\{\ln \tilde{X}(u+(k-1))\right\}_{k\in\mathbb{N}_+}$ is stationary and ergodic. Further, $\left(\ln \tilde{X}(u+(k-1))\right)^2$ is a measurable function of the stationary and ergodic process $\left\{\ln \tilde{X}(u+(k-1))\right\}_{i\in\mathbb{N}_+}$, by Theorem 3.5.8 in Stout [1974], $\left\{(\ln \tilde{X}(u+(k-1)))^2\right\}_{k\in\mathbb{N}_+}$ is stationary and ergodic. So *the point wise ergodic theorem for stationary sequences* (Theorem 3.5.7 in Stout [1974] can be applied to the sequence $\left\{\int_{j-1}^{j} (\ln \tilde{X}(t))^2 dt\right\}_{j\in\mathbb{N}_+}$. Thus,

$$\text{LHS of } (A.76) \xrightarrow[n\to\infty]{a.s.\text{ and }L^{m/2}} \int_0^1 (\tilde{r}^2(u) + \frac{\sigma^2}{2\alpha})du = \int_0^1 (\tilde{r}(t))^2 dt + \frac{\sigma^2}{2\alpha}.$$

Similarly,

$$\begin{aligned}\text{RHS of } (A.76) &= \frac{1}{T}\int_0^{\lfloor T\rfloor+1}(\ln\tilde{X}(t))^2dt \xrightarrow[n\to\infty]{a.s. \text{ and } L^{m/2}} \int_0^1(\tilde{r}^2(u)+\frac{\sigma^2}{2\alpha})du\\ &= \int_0^1(\tilde{r}(t))^2dt+\frac{\sigma^2}{2\alpha}.\end{aligned}$$

This implies that

$$\frac{1}{T}\int_0^T(\ln\tilde{X}(t))^2dt \xrightarrow[T\to\infty]{a.s. \text{ and } L^{m/2}} \int_0^1(\tilde{r}(t))^2dt+\frac{\sigma^2}{2\alpha}.$$

(4) From Proposition 2.3.1 and the fact that Σ is invertible, then, continuous mapping theorem can be applied to get $TQ^{-1}_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \Sigma^{-1}$.

Step **(1)**,**(2)**, **(3)**, and **(4)** complete the proof. □

Proof of Proposition 2.3.5. This proposition is a special case of Proposition 1.21 in *Kutoyants* Kutoyants [2004], for which $d_1 = p+1$ and $d_2 = 1$. As defined in (2.3.3)

$$W_{[0,T]} = \left[\int_0^T\varphi_1(t)dB_t,\ \int_0^T\varphi_2(t)dB_t,\cdots,\int_0^T\varphi_p(t)dB_t,-\int_0^T\ln X(t)dB_t\right]^\top$$

which is a $p+1$ column vector. Let $h_T^{(i)}(t,\omega) = \frac{1}{\sqrt{T}}\varphi_i(t)$ for $i = 1,2,\cdots,p$ and $h_T^{(p+1)}(t,\omega) = \frac{1}{\sqrt{T}}\ln X(,\omega)$. This implies that $\left(h_T^{(i)}(t,\omega)\right)^2 = \frac{1}{T}\varphi_i^2(t)$ and $\left(h_T^{(p+1)}(t,\omega)\right)^2 = \frac{1}{T}(\ln X(t,\omega))^2$, for all $i = 1,2,\cdots,p$. By (A.55),

$$\sup_{t\geq 0}\mathbb{E}\left[|\ln X(t)|^m\right] \leq 3^{m-1}\left(\mathbb{E}\left[|\ln X_0|^m\right]+\left(\frac{K_\mu K_\varphi+\frac{1}{2}\sigma^2}{\alpha}\right)^m+C_{m/2}\sigma^m\left(\frac{1}{2\alpha}\right)^{m/2}\right)<\infty,$$

and $|\varphi_i|(t) \leq K_\varphi < \infty$, which implies that $\mathrm{P}\left(\int_0^T\left(h_T^{(i)}(t,\omega)\right)^2dt<\infty\right) = 1$ for all $i = 1,2,\cdots,p+1$. Further,

$$\int_0^T\frac{1}{\sqrt{T}}\varphi^\top(t)\frac{1}{\sqrt{T}}\varphi(t)dt = \frac{1}{T}\int_0^T\varphi^\top(t)\varphi(t)dt,$$

$$\int_0^T\frac{1}{\sqrt{T}}\varphi(t)\frac{1}{\sqrt{T}}\ln X(t)dt = \frac{1}{T}\int_0^T\varphi(t)\ln X(t)dt,$$

and

$$\int_0^T \frac{1}{\sqrt{T}} \ln X(t) \frac{1}{\sqrt{T}} \ln X(t) dt = \frac{1}{T} \int_0^T (\ln X(t))^2 \, dt.$$

Then, by (2.3.2),

$$\frac{1}{T} Q_{[0,T]} = \begin{bmatrix} \frac{1}{T} \int_0^T \varphi^\top(t)\varphi(t)dt & -\frac{1}{T} \int_0^T \varphi^\top(t) \ln X(t)dt \\ -\frac{1}{T} \int_0^T \varphi(t) \ln X(t)dt & \frac{1}{T} \int_0^T (\ln X(t))^2 \, dt \end{bmatrix}$$

and by Proposition 2.3.4, $\frac{1}{T} Q_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \Sigma$. Finally, Proposition 1.21 in Kutoyants [2004] implies that

$$\frac{1}{\sqrt{T}} W_{[0,T]} \xrightarrow[T\to\infty]{D} W^* \sim \mathcal{N}_{p+1}(0, \Sigma).$$

This completes the proof. □

Proof of Proposition 2.3.6. From (2.3.5), $\hat{\theta}_T = \theta + \sigma Q_{[0,T]}^{-1} W_{[0,T]}$. Proposition 2.3.4 shows that $TQ_{[0,T]}^{-1} \xrightarrow[T\to\infty]{a.s.} \Sigma^{-1}$, and further, the proof of L^2 boundedness of $\frac{1}{T} W_{[0,T]}$ is given. $W_{[0,T]}$ is defined as

$$W_{[0,T]} = \left[\int_0^T \varphi_1(t)dB_t, \int_0^T \varphi_2(t)dB_t, \cdots, \int_0^T \varphi_p(t)dB_t, -\int_0^T \ln X(t)dB_t \right]^\top.$$

For $i = 1, \cdots, p$, $\int_0^T \varphi_1(t)dB_t$ is a martingale, and

$$\sup_{T\geq 0} \mathbb{E}\left[\left| \frac{1}{\sqrt{T}} \int_0^T \varphi_i(t)dB_t \right|^2 \right] = \sup_{T\geq 0} \mathbb{E}\left[\frac{1}{T} \int_0^T \varphi_i^2(t)dt \right] \leq K_\varphi^2,$$

and from (A.55), with $m = 2$,

$$\sup_{T\geq 0} \mathbb{E}\left[\left| \frac{1}{\sqrt{T}} \int_0^T \ln X(t)(t)dB_t \right|^2 \right] = \sup_{T\geq 0} \mathbb{E}\left[\frac{1}{T} \int_0^T (\ln X(t))^2 dt \right] \leq \sup_{t\geq 0} \mathbb{E}\left[(\ln X(t))^2 \right] < \infty,$$

which completes the L^2 boundedness. Further, $W_{[0,T]}$ is a martingale, by using Doob's maximal inequality for submartingales, for any $\varepsilon > 0$,

$$\mathrm{P}\left(\sup_{2^k \leq T \leq 2^{k+1}} \frac{1}{T} |W_{[0,T]}| > \varepsilon \right) \leq \mathrm{P}\left(\sup_{2^k \leq T \leq 2^{k+1}} |W_{[0,T]}| > \varepsilon 2^k \right)$$

$$\leq \frac{\mathbb{E}\left[\sup\limits_{2^k\leq T\leq 2^{k+1}}|W_{[0,T]}|^2\right]}{\varepsilon^2 2^{2k}} \leq \frac{\mathbb{E}\left[|W_{[0,2^{k+1}]}|^2\right]}{\varepsilon^2 2^{2k}} \leq \frac{\max\left\{\sup\limits_{t\geq 0}\mathbb{E}\left[(\ln X(t))^2\right], K_\varphi^2\right\}2^{k+1}}{\varepsilon^2 2^{2k}} = O\left(\frac{1}{2^k}\right).$$

$\frac{1}{T}|W_{[0,T]}| \xrightarrow[T\to\infty]{a.s.} 0$ can be obtained by applying the Borel–Cantelli Lemma. Then,$\hat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$. Further, from (2.2.1), $dX(t) = S(\theta, t, X(t))dt + \sigma X(t)dB_t$, then, Proposition 2.3.2 implies that $\hat{\theta}_T = \theta + \sigma Q^{-1}_{[0,T]}W_{[0,T]}$. By the definition of ρ_T,

$$\sqrt{T}(\hat{\theta}_T - \theta) = \sigma\sqrt{T}Q^{-1}_{[0,T]}W_{[0,T]} = \sigma T Q^{-1}_{[0,T]}\frac{1}{\sqrt{T}}W_{[0,T]}$$

By Proposition 2.3.4, $\sigma T Q^{-1}_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \sigma\Sigma^{-1}$. Proposition 2.3.5 stats

$$\frac{1}{\sqrt{T}}W_{[0,T]} \xrightarrow[T\to\infty]{D} W^* \sim \mathcal{N}_{p+1}(0,\Sigma). \tag{A.77}$$

Then, by Slutsky's Theorem,

$$\sqrt{T}(\hat{\theta}_T - \theta) = \sigma T Q^{-1}_{[0,T]}\frac{1}{\sqrt{T}}W_{[0,T]} \xrightarrow[T\to\infty]{D} \sigma\Sigma^{-1}W^* = \rho.$$

Note that Σ^{-1} is non-random and symmetric, by the proposition of multivariate normal distribution, $\rho \sim \mathcal{N}_{p+1}\left(0, \sigma^2\Sigma^{-1}\right)$. Then,

$$\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow[T\to\infty]{D} \rho \sim \mathcal{N}_{p+1}\left(0, \sigma^2\Sigma^{-1}\right).$$

This completes the proof. □

Proof of Proposition 2.3.7. From (2.3.8),

$$\sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) = \sqrt{T}(\hat{\theta}_T - \theta) - \sqrt{T}(\tilde{\theta}_T - \theta) = G_{[0,T]}M\sqrt{T}(\hat{\theta}_T - \theta) + \sqrt{T}G_{[0,T]}(M\theta - r).$$

Hence,

$$\begin{bmatrix}\sqrt{T}(\hat{\theta}_T - \theta)\\ \sqrt{T}(\tilde{\theta}_T - \theta)\\ \sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T)\end{bmatrix} = \begin{bmatrix} I_{p+1}\\ I_{p+1} - G_{[0,T]}M\\ G_{[0,T]}M\end{bmatrix}\sqrt{T}(\hat{\theta}_T - \theta) + \begin{bmatrix}0\\ -\sqrt{T}G_{[0,T]}(M\theta - r)\\ \sqrt{T}G_{[0,T]}(M\theta - r)\end{bmatrix}.$$

By the fact

$$G_{[0,T]} \xrightarrow[T\to\infty]{\mathrm{P}} G^* = \Sigma^{-1}M^\top(M\Sigma^{-1}M^\top)^{-1}, \tag{A.78}$$

$$\begin{bmatrix} I_{p+1} \\ I_{p+1} - G_{[0,T]}M \\ G_{[0,T]}M \end{bmatrix} \xrightarrow[T\to\infty]{\mathrm{P}} \begin{bmatrix} I_{p+1} \\ I_{p+1} - G^{*}M \\ G^{*}M \end{bmatrix}.$$

Consider the local alternatives restriction (2.3.7),

$$\sqrt{T}G_{[0,T]}(M\theta - r) = \sqrt{T}G_{[0,T]}\frac{r_0}{\sqrt{T}} = G_{[0,T]}r_0 \xrightarrow[T\to\infty]{\mathrm{P}} G^{*}r_0.$$

Then,

$$\begin{bmatrix} 0 \\ -\sqrt{T}G_{[0,T]}(M\theta - r) \\ \sqrt{T}G_{[0,T]}(M\theta - r) \end{bmatrix} \xrightarrow[T\to\infty]{\mathrm{P}} \begin{bmatrix} 0 \\ -G^{*}r_0 \\ G^{*}r_0 \end{bmatrix}.$$

By Proposition 2.3.6, and Slutsky's Theorem, that

$$\begin{bmatrix} \sqrt{T}(\hat{\theta}_T - \theta) \\ \sqrt{T}(\tilde{\theta}_T - \theta) \\ \sqrt{T}(\hat{\theta}_T - \tilde{\theta}_T) \end{bmatrix} \xrightarrow[T\to\infty]{D} \begin{bmatrix} I_{p+1} \\ I_{p+1} - G^{*}M \\ G^{*}M \end{bmatrix} \rho + \begin{bmatrix} 0 \\ -G^{*}r_0 \\ G^{*}r_0 \end{bmatrix} = \begin{bmatrix} \rho \\ \varrho \\ \varsigma \end{bmatrix}.$$

Note that the mean of $(\rho^{\top}, \varrho^{\top}, \varsigma^{\top})^{\top}$ is

$$\begin{bmatrix} I_{p+1} \\ I_{p+1} - G^{*}M \\ G^{*}M \end{bmatrix} 0 + \begin{bmatrix} 0 \\ -G^{*}r_0 \\ G^{*}r_0 \end{bmatrix} = \begin{bmatrix} 0 \\ -G^{*}r_0 \\ G^{*}r_0 \end{bmatrix}$$

The variance of $(\rho^{\top}, \varrho^{\top}, \varsigma^{\top})^{\top}$ is

$$\begin{bmatrix} I_{p+1} \\ I_{p+1} - G^{*}M \\ G^{*}M \end{bmatrix} \sigma^2\Sigma^{-1} \begin{bmatrix} I_{p+1} \\ I_{p+1} - G^{*}M \\ G^{*}M \end{bmatrix}^{\top} = \sigma^2 \begin{bmatrix} \Sigma^{-1} \\ \Sigma^{-1} - G^{*}M\Sigma^{-1} \\ G^{*}M\Sigma^{-1} \end{bmatrix} \begin{bmatrix} I_{p+1} \\ I_{p+1} - G^{*}M \\ G^{*}M \end{bmatrix}^{\top}$$

$$= \sigma^2 \begin{bmatrix} \Sigma^{-1} & \Sigma^{-1} - \Sigma^{-1}(G^*M)^\top & \Sigma^{-1}(G^*M)^\top \\ \Sigma^{-1} - G^*M\Sigma^{-1} & (\Sigma^{-1} - G^*M\Sigma^{-1})(I_{p+1} - (G^*M)^\top) & (\Sigma^{-1} - G^*M\Sigma^{-1})(G^*M)^\top \\ G^*M\Sigma^{-1} & G^*M\Sigma^{-1}(I_{p+1} - (G^*M)^\top) & G^*M\Sigma^{-1}(G^*M)^\top \end{bmatrix}.$$

From (A.78),

$$\Sigma^{-1}(G^*M)^\top = \Sigma^{-1}M^\top(M\Sigma^{-1}M^\top)^{-1}M\Sigma^{-1} = G^*M\Sigma^{-1},$$

and

$$G^*M\Sigma^{-1}(G^*M)^\top = \left(\Sigma^{-1}M^\top(M\Sigma^{-1}M^\top)^{-1}\right)M\Sigma^{-1}\left(M^\top(M\Sigma^{-1}M^\top)^{-1}M\Sigma^{-1}\right)$$

$$= \Sigma^{-1}M^\top(M\Sigma^{-1}M^\top)^{-1}(M\Sigma^{-1}M^\top)(M\Sigma^{-1}M^\top)^{-1}M\Sigma^{-1} = G^*M\Sigma^{-1}.$$

As what has been shown, the conclusion

$$\begin{bmatrix} \rho_T \\ \varrho_T \\ \varsigma_T \end{bmatrix} \xrightarrow[T\to\infty]{\mathrm{D}} \begin{bmatrix} \rho \\ \varrho \\ \varsigma \end{bmatrix} \sim \mathcal{N}_{3(p+1)}\left[\begin{bmatrix} 0 \\ -G^*r_0 \\ G^*r_0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \Sigma^{-1} & \Sigma^{-1} - G^*M\Sigma^{-1} & G^*M\Sigma^{-1} \\ \Sigma^{-1} - G^*M\Sigma^{-1} & \Sigma^{-1} - G^*M\Sigma^{-1} & 0 \\ G^*M\Sigma^{-1} & 0 & G^*M\Sigma^{-1} \end{bmatrix}\right].$$

This finishes the proof of the joint asymptotic normality of the UMLE and RMLE. This completes the proof. □

Proof of Proposition 2.3.8. From Proposition 2.3.7,

$$\varsigma_T \xrightarrow[T\to\infty]{D} \varsigma \sim \mathcal{N}_{p+1}(G^*r_0, \sigma^2 G^*M\Sigma^{-1}).$$

Further, by Proposition 2.3.4,

$$\hat{\Gamma} = \frac{1}{\hat{\sigma}^2}M^\top\left(MTQ_{[0,T]}^{-1}M^\top\right)^{-1}M \xrightarrow[T\to\infty]{\mathrm{P}} \Gamma = \frac{1}{\sigma^2}M^\top\left(M\Sigma^{-1}M^\top\right)^{-1}M.$$

Therefore, by Slutsky's Theorem, $\psi_T = \varsigma_T^\top\hat{\Gamma}\varsigma_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi = \varsigma^\top\Gamma\varsigma$. To complete the proof of this proposition, it suffices to apply Theorem 5.1.3 in Mathai and Provost [1992] along with some algebraic computations. □

Proof of Proposition 2.3.11. By (2.3.16),

$$\begin{aligned}\mathbb{E}\left[|\ln X_1(t)|^m\right] &= \mathbb{E}\left[\left|e^{-\alpha_1 t}\ln X_0 + e^{-\alpha_1 t}\int_0^t e^{\alpha_1 s}(L(s)-\tfrac{1}{2}\sigma^2)ds + \sigma e^{-\alpha_1 t}\int_0^t e^{\alpha_1 s}dB_s\right|^m\right]\\ &\leqslant 3^{m-1}\left(e^{-m\alpha_1 t}\mathbb{E}\left[|\ln X_0|^m\right] + \left|e^{-\alpha_1 t}\int_0^t e^{\alpha_1 s}(L(s)-\tfrac{1}{2}\sigma^2)ds\right|^m\right)\\ &+\sigma^m 3^{m-1}e^{-m\alpha_1 t}\mathbb{E}\left[\left|\int_0^t e^{\alpha_1 s}dB_s\right|^m\right],\end{aligned}$$

and

$$\left|e^{-\alpha_1 t}\int_0^t e^{\alpha_1 s}(L(s)-\frac{1}{2}\sigma^2)ds\right|^m \leqslant \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha_1}\right)^m (1-e^{-\alpha_1 t})^m.$$

Further, by Burkholder-Davis-Gundy inequality, there exists a positive constant $C_{m/2}$, such that

$$\sigma^m e^{-m\alpha_1 t}\mathbb{E}\left[\left|\int_0^t e^{\alpha_1 s}dB_s\right|^m\right] \leqslant \sigma^m e^{-m\alpha_1 t}\mathbb{E}\left[\left(\sup_{0\leqslant s\leqslant t}\left|\int_0^s e^{\alpha_1 s}dB_s\right|\right)^m\right]$$

$$\leqslant C_{m/2}\sigma^m e^{-m\alpha_1 t}\mathbb{E}\left[\left(\int_0^t e^{2\alpha_1 s}ds\right)^{m/2}\right] = C_{m/2}\sigma^m e^{-m\alpha_1 t}\left(\frac{1}{2\alpha_1}(e^{2\alpha_1 t}-1)\right)^{m/2},$$

$$= C_{m/2}\sigma^m\left(\frac{1}{2\alpha_1}(1-e^{-2\alpha_1 t})\right)^{m/2},$$

and then, since $\mathbb{E}\left[|\ln X_0|^m\right] < \infty$,

$$\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X_1(t)|^m\right] \leqslant 3^{m-1}\left(\mathbb{E}\left[|\ln X_0|^m\right] + \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha_1}\right)^m + C_{m/2}\sigma^m\left(\frac{1}{2\alpha_1}\right)^{m/2}\right) < \infty. \tag{A.79}$$

Furthermore, by (2.3.16),

$$\mathbb{E}\left[|\ln X_2(t)|^m\right] \leqslant 3^{m-1}\left(|e^{-\alpha_2(t-t^*)}\ln X_0^{t^*})^m + (r_2^{t^*}(t-t^*))^m + (\tau_2^{t^*}(t-t^*)|^m\right). \tag{A.80}$$

First, notice that $\ln X^{t^*}(t) = \ln X(t+t^*)$, which implies that $\ln X^{t^*}(0) = \ln X(t^*)$. From (A.79), $|e^{-\alpha_2(t-t^*)}\ln X^{t^*}(0)|^m \leqslant |\ln X^{t^*}(0)|^m < \infty$, for all $t \geqslant t^*$. Second,

$$\begin{aligned}&\left|r_2^{t^*}(t-t^*)\right|^m \leqslant e^{-m\alpha_2(t-t^*)}\left(\int_0^{(t-t^*)} e^{\alpha_2 s}\left|L(s+t^*)-\frac{1}{2}\sigma^2\right|ds\right)^m\\ &\leqslant \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha_2}\right)^m (1-e^{-\alpha_2(t-t^*)})^m,\end{aligned} \tag{A.81}$$

which is bounded in t on $[t^*, +\infty)$. Further, by Burkholder-Davis-Gundy inequality, there exists a positive constant $C_{m/2}$, such that

$$\begin{aligned}\left|\tau_2^{t^*}(t-t^*)\right|^m &= \sigma^m e^{-m\alpha_2(t-t^*)}\mathbb{E}\left[\left|\int_0^{(t-t^*)} e^{\alpha_1 s} dB_s^{t^*}\right|^m\right] \\ &\leqslant C_{m/2}\sigma^m e^{-m\alpha_2(t-t^*)}\mathbb{E}\left[\left(\int_0^{(t-t^*)} e^{2\alpha_2 s} ds\right)^{m/2}\right],\end{aligned} \tag{A.82}$$

and then, by (A.80), this implies that

$$\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X_2(t)|^m\right] \leqslant 3^{m-1}\left(\mathbb{E}\left[|\ln X^{t^*}(0)|^m\right] + \left(\frac{K_\mu K_\varphi + \frac{1}{2}\sigma^2}{\alpha_2}\right)^m + C_{m/2}\sigma^m\left(\frac{1}{2\alpha_2}\right)^{m/2}\right) < \infty. \tag{A.83}$$

The rest of the proof follows from (A.79) and (A.83). □

Proof of Proposition 2.3.15. From Proposition A.18 and Theorem 7.6 in Liptser and Shiryaev [2001], the likelihood function of the diffusion process (2.3.12) is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}, s, X^T) &:= \frac{dP_X}{dP_B}(X^T) = \exp\left\{\int_0^T \frac{S(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)}\mathbb{I}_{\{0\leqslant t\leqslant t^*\}}dX(t)\right. \\ &+\int_0^T \frac{S(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)}\mathbb{I}_{\{t^*<t\leqslant T\}}dX(t) \\ &\left.-\frac{1}{2}\int_0^T \frac{S^2(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)}\mathbb{I}_{\{0\leqslant t\leqslant t^*\}}dt - \frac{1}{2}\int_0^T \frac{S^2(\boldsymbol{\theta}, t, X(t))}{\sigma^2 X^2(t)}\mathbb{I}_{\{t^*<t\leqslant T\}}dt\right\}.\end{aligned} \tag{A.84}$$

$$\begin{aligned}= \exp\Bigg\{&\int_0^{sT} \frac{(\varphi(t), -\ln X(t))\theta^{(1)}X(t)}{\sigma^2 X^2(t)}dX(t) + \int_{sT}^T \frac{(\varphi(t), -\ln X(t))\theta^{(2)}X(t)}{\sigma^2 X^2(t)}dX(t) \\ &-\frac{1}{2}\int_0^{sT} \frac{\theta^{(1)\top}(\varphi(t), -\ln X(t))^\top(\varphi(t), -\ln X(t))\theta^{(1)}X(t)X(t)}{\sigma^2 X^2(t)}dt \\ &-\frac{1}{2}\int_{sT}^T \frac{\theta^{(2)\top}(\varphi(t), -\ln X(t))^\top(\varphi(t), -\ln X(t))\theta^{(2)}X(t)X(t)}{\sigma^2 X^2(t)}dt\Bigg\}.\end{aligned}$$

This gives

$$\begin{aligned}\log\mathcal{L}(\boldsymbol{\theta}, s, X^T) &= \frac{1}{\sigma^2}\int_0^{sT} \frac{(\varphi(t), -\ln X(t))}{X(t)}dX(t)\theta^{(1)} + \frac{1}{\sigma^2}\int_{sT}^T \frac{(\varphi(t), -\ln X(t))}{X(t)}dX(t)\theta^{(2)} \\ &-\frac{1}{2\sigma^2}\theta^{(1)\top}\int_0^{sT}(\varphi(t), -\ln X(t))^\top(\varphi(t), -\ln X(t))dt\theta^{(1)} \\ &-\frac{1}{2\sigma^2}\theta^{(2)\top}\int_{sT}^T(\varphi(t), -\ln X(t))^\top(\varphi(t), -\ln X(t))dt\theta^{(2)} = \frac{1}{\sigma^2}U^\top(s,T)\boldsymbol{\theta} - \frac{1}{2\sigma^2}\boldsymbol{\theta}^\top Q(s,T)\boldsymbol{\theta}.\end{aligned}$$

First, take the first derivative of $\log\mathcal{L}(\boldsymbol{\theta}, s, X^T)$ with respect to $\boldsymbol{\theta}$.

$$\frac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{L}(\boldsymbol{\theta}, s, X^T) = \frac{1}{\sigma^2}U^\top(s,T) - \frac{1}{\sigma^2}\boldsymbol{\theta}^\top Q(s,T).$$

Set $\frac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{L}(\boldsymbol{\theta}, s, X^T) = 0$, $\hat{\theta}_T(s) = Q^{-1}(s,T)U(s,T)$. From Proposition 2.3.14, $Q(s,T)$ is positive definite, which implies that $Q(s,T)$ is invertable. Next, taking the second derivative of $\log\mathcal{L}(\boldsymbol{\theta}, s, X^T)$ with respect to $\boldsymbol{\theta}^\top$ gives,

$$\frac{\partial^2}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\log\mathcal{L}(\boldsymbol{\theta}, s, X^T) = -\frac{1}{\sigma^2}Q(s,T).$$

By Proposition 2.3.14, $-\frac{1}{\sigma^2}Q(s,T)$ is a negative definite matrix, since $\sigma^2 > 0$. This implies that $Q^{-1}(s,T)U(s,T)$ is the maximum value, and $\hat{\theta}_T(s) = Q^{-1}(s,T)U(s,T)$. Further, the maximum value of the log likelihood function $\log\mathcal{L}(\boldsymbol{\theta}, s, X^T)$ under the restriction (2.3.14) can be found. Here Lagrange multiplier method is used. After introducing the Lagrange multiplier into the log likelihood function,

$$\log\mathcal{L}(\boldsymbol{\theta}, s, X^T, \lambda) = \frac{1}{\sigma^2}U^\top(s,T)\boldsymbol{\theta} - \frac{1}{2\sigma^2}\boldsymbol{\theta}^\top Q(s,T)\boldsymbol{\theta} + \lambda^\top(M\boldsymbol{\theta} - r)$$

where λ is a q column vector. To find the maximum value point of $\log\mathcal{L}(\boldsymbol{\theta}, s, X^T, \lambda)$, taking the first derivative with respect to $\boldsymbol{\theta}$ and λ respectively are needed,

$$\frac{\partial}{\partial\boldsymbol{\theta}}\log\mathcal{L}(\boldsymbol{\theta}, s, X^T, \lambda) = \frac{1}{\sigma^2}U^\top(s,T) - \frac{1}{\sigma^2}\boldsymbol{\theta}^\top Q(s,T) + \tilde{M}^\top\lambda$$

and

$$\frac{\partial}{\partial\lambda}\log\mathcal{L}(\boldsymbol{\theta}, s, X^T, \lambda) = \tilde{M}\boldsymbol{\theta} - \tilde{r}.$$

By setting the first partial derivatives equal to 0,

$$\frac{1}{\sigma^2}U^\top(s,T) - \frac{1}{\sigma^2}\boldsymbol{\theta}^\top Q(s,T) + \tilde{M}^\top\lambda = 0 \tag{A.85}$$

and

$$\tilde{M}\boldsymbol{\theta} - \tilde{r} = 0. \tag{A.86}$$

$\frac{1}{\sigma^2}U^\top(s,T) + \tilde{M}^\top\lambda = \frac{1}{\sigma^2}\boldsymbol{\theta}^\top Q(s,T)$, and this implies that

$$\tilde{\theta}_T(s) = Q^{-1}(s,T)U(s,T) + \sigma^2 Q^{-1}(s,T)\tilde{M}^\top\lambda. \tag{A.87}$$

Substituting (A.85) into (A.86), $\tilde{M}Q^{-1}(s,T)U(s,T) + \tilde{M}\sigma^2 Q^{-1}(s,T)\tilde{M}^\top\lambda = \tilde{r}$. Then, $\tilde{M}\sigma^2 Q^{-1}(s,T)\tilde{M}^\top\lambda = \tilde{r} - \tilde{M}Q^{-1}(s,T)U(s,T)$,

$$\hat{\lambda} = \frac{1}{\sigma^2}(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top)^{-1}\tilde{r} - \frac{1}{\sigma^2}(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top)^{-1}\tilde{M}Q^{-1}(s,T)U(s,T).$$

Substituting $\hat{\lambda}$ into (A.87),

$$\tilde{\theta}_T(s) = Q^{-1}(s,T)U(s,T) + \sigma^2 Q^{-1}(s,T)\tilde{M}^\top\left(\frac{1}{\sigma^2}\left(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top\right)^{-1}\tilde{r}\right.$$

$$\left.-\frac{1}{\sigma^2}\left(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top\right)^{-1}\tilde{M}Q^{-1}(s,T)U(s,T)\right).$$

Then,

$$\tilde{\theta}_T(s) = Q^{-1}(s,T)U(s,T) + Q^{-1}(s,T)\tilde{M}^\top(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top)^{-1}\tilde{r}$$

$$-Q^{-1}(s,T)\tilde{M}^\top(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top)^{-1}\tilde{M}Q^{-1}(s,T)U(s,T).$$

This gives $\tilde{\theta}_T(s) = \hat{\theta}_T(s) + \tilde{G}_{[0,T]}\tilde{r} - \tilde{G}_{[0,T]}\tilde{M}\hat{\theta}_T(s)$, where $\hat{\theta}_T(s) = Q^{-1}(s,T)U(s,T)$ and $\tilde{G}_{[0,T]} = Q^{-1}(s,T)\tilde{M}^\top(\tilde{M}Q^{-1}(s,T)\tilde{M}^\top)^{-1}$. Finally,

$$\tilde{\theta}_T(s) = \hat{\theta}_T(s) + \tilde{G}_{[0,T]}\tilde{r} - \tilde{G}_{[0,T]}\tilde{M}\hat{\theta}_T(s) = \hat{\theta}_T(s) - \tilde{G}_{[0,T]}(\tilde{M}\hat{\theta}_T(s) - \tilde{r}).$$

This completes the proof. □

Lemma A.6. *Let $\hat{s}$ be $\mathscr{F}_T$−measurable and a consistent estimator of s, with $0 \leqslant \hat{s} \leqslant 1$ a.s. and let $\{\varphi(t), t \geqslant 0\}$ be the deterministic and bounded function. Then,*

$$\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}\varphi(t)dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}\varphi(t)dB_t \xrightarrow[T\to\infty]{L^m} 0.$$

Proof. First, the following convergence needs to be derived

$$\lim_{T\to\infty} T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m\right] = 0. \tag{A.88}$$

Let $\|\varphi(t)\| \leqslant K_\varphi$ and $\epsilon > 0$,

$$\lim_{T\to\infty} \mathrm{P}\left(|\hat{s} - s| \geqslant \epsilon/(4K_\varphi^2)\right) = 0. \tag{A.89}$$

Let $G(T) = T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m\right]$, further, let

$$\begin{aligned}
G_{11}(T) &= T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m \mathbb{I}_{\{\hat{s}>s\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\frac{\epsilon}{4K_\varphi^2}\}}\right]\\
G_{12}(T) &= T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m \mathbb{I}_{\{\hat{s}>s\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\frac{\epsilon}{4K_\varphi^2}\}}\right]\\
G_{21}(T) &= T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m \mathbb{I}_{\{\hat{s}<s\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\frac{\epsilon}{4K_\varphi^2}\}}\right]\\
G_{22}(T) &= T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m \mathbb{I}_{\{\hat{s}<s\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\frac{\epsilon}{4K_\varphi^2}\}}\right].
\end{aligned} \tag{A.90}$$

Note that (A.88) is equivalent to $\lim_{T\to\infty}(G_{11}(T) + G_{12}(T) + G_{21}(T) + G_{22}(T)) = 0$. For the convergence of $G_{11}(T)$, $G_{11}(T) = T^{-m/2}\mathbb{E}\left[\left|\int_{sT}^{\hat{s}T}\varphi(t)dB_t\right|^m \mathbb{I}_{\{0<(\hat{s}-s)T\leqslant\frac{\epsilon T}{4K_\varphi^2}\}}\right]$. Let $u = t - sT$ with $u \in [0, (\hat{s} - s)T]$, and $dt = du$,

$$\begin{aligned}
G_{11}(T) &= T^{-m/2}\mathbb{E}\left[\left|\int_0^{(\hat{s}-s)T}\varphi(u + sT)dB_{u+sT}\right|^m \mathbb{I}_{\{0<(\hat{s}-s)T\leqslant\frac{\epsilon T}{4K_\varphi^2}\}}\right]\\
&\leqslant T^{-m/2}\mathbb{E}\left[\sup_{0\leqslant t\leqslant\frac{\epsilon T}{4K_\varphi^2}}\left|\int_0^t\varphi(u + sT)dB_{u+sT}\right|^m\right].
\end{aligned}$$

Further, by Burkholder-Davis-Gundy inequality, there exists some positive constant $C_{m/2}$, such that

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant\frac{\epsilon T}{4K_\varphi^2}}\left|\int_0^t\varphi(u + sT)dB_{u+sT}\right|^m\right] \leqslant C_{m/2}\mathbb{E}\left[\left|\int_0^{\frac{\epsilon T}{4K_\varphi^2}}\varphi^2(u + sT)du\right|^{m/2}\right] \leqslant C_{m/2}\left(\frac{\epsilon T}{4}\right)^{m/2},$$

which implies that $G_{11}(T) \leqslant T^{-m/2}C_{m/2}\left(\frac{\epsilon T}{4}\right)^{m/2} = C_{m/2}\left(\frac{1}{4}\right)^{m/2}\epsilon^{m/2}$. Further, as $0 < s < 1$ and $0 < \hat{s} < 1$ a.s.,

$$G_{12}(T) = T^{-m/2}\mathbb{E}\left[\left|\int_0^{\hat{s}T}\varphi(t)dB_t - \int_0^{sT}\varphi(t)dB_t\right|^m \mathbb{I}_{\{1>\hat{s}-s>0\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\frac{\epsilon}{4K_\varphi^2}\}}\right]$$

$$\leqslant T^{-m/2}\mathbb{E}\left[\left|\int_{sT}^{\hat{s}T}\varphi(t)dB_t\right|^m \mathbb{I}_{\{T>(\hat{s}-s)T>0\}}\mathbb{I}_{\{\hat{s}-s\geqslant\frac{\epsilon}{4K_\varphi^2}\}}\right].$$

Let $u = t - sT$, then, $u \in [0, (\hat{s} - s)T]$ and $dt = du$,

$$G_{12}(T) \leqslant T^{-m/2}\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|\int_0^t \varphi(u+sT)dB_{u+sT}\right|^m \mathbb{I}_{\{T>(\hat{s}-s)T>0\}}\mathbb{I}_{\{\hat{s}-s\geqslant\frac{\epsilon}{4K_\varphi^2}\}}\right]$$

$$\leqslant T^{-m/2}\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|\int_0^t \varphi(u+sT)dB_{u+sT}\right|^m \mathbb{I}_{\{T>(\hat{s}-s)T\geqslant\frac{\epsilon T}{4K_\varphi^2}\}}\right].$$

Furthermore, by Cauchy-Schwartz inequality,

$$G_{12}(T) \leqslant T^{-m/2}\left\{\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|\int_0^t \varphi(u+sT)dB_{u+sT}\right|^{2m}\right]\right\}^{1/2}\{\mathbb{E}[\mathbb{I}_{\{(\hat{s}-s)T\geqslant\frac{\epsilon T}{4K_\varphi^2}\}}]\}^{1/2}$$

$$\leqslant T^{-m/2}\left\{\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|\int_0^t \varphi(u+sT)dB_{u+sT}\right|^{2m}\right]\right\}^{1/2}\left\{\mathrm{P}\left(\hat{s}-s\geqslant\frac{\epsilon}{4K_\varphi}\right)\right\}^{1/2}.$$

Then, by Burkholder-Davis-Gundy's inequality, there exists a C_m, such that

$$\mathbb{E}\left[\sup_{0\leqslant t\leqslant T}\left|\int_0^t \varphi(u+sT)dB_{u+sT}\right|^{2m}\right] \leqslant C_m\mathbb{E}\left[\int_0^T |\varphi(u)|^2 du\right]^m \leqslant C_m(K_\varphi^2 T)^m. \tag{A.91}$$

Then,

$$G_{12}(T) \leqslant T^{-m/2}\left\{C_m(K_\varphi^2 T)^m\right\}^{1/2}\left\{\mathrm{P}\left(\hat{s}-s\geqslant\tfrac{\epsilon}{4K_\varphi^2}\right)\right\}^{1/2} \leqslant \sqrt{C_m}(K_\varphi)^m\left\{\mathrm{P}\left(\hat{s}-s\geqslant\tfrac{\epsilon}{4K_\varphi^2}\right)\right\}^{1/2}.$$

Together with the fact of (A.89), that $\lim_{T\to\infty} G_{12}(T) = 0$. $\lim_{T\to\infty} G_{21}(T) = \lim_{T\to\infty} G_{22}(T) = 0$ can be proven by using the similar techniques as for $G_{11}(T)$ and $G_{12}(T)$. This implies that (A.88) holds and completes the proof. □

Lemma A.7. *Suppose that Assumption 2.1-2.3, and Assumption 2.4 hold and let $\hat{s}$ be a consistent estimator of s, with $0 < \hat{s} < 1$ a.s. Then,*

$$\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t \xrightarrow[T\to\infty]{L^{m/2}} 0.$$

Proof. Since $\hat{s}$ is $\mathscr{F}_T$−measurable and a consistent estimator of s, let $\epsilon > 0$, for every

$0 < \delta < \min\{s, 1-s\}$, $\mathrm{P}(|\hat{s}-s| \geqslant \delta) < \epsilon$, for sufficiently large T. Then,

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\right]$$

$$= \mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{|\hat{s}-s|\geqslant\delta\}}\right] \tag{A.92}$$

$$+\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{|\hat{s}-s|<\delta\}}\right]. \tag{A.93}$$

$$\begin{aligned}(A.92) = {} & \mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{\hat{s}\leqslant s-\delta\}}\right] \\ & +\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{\hat{s}\geqslant s+\delta\}}\right] \\ \leqslant {} & T^{-m/4}\mathbb{E}\left[\sup_{0\leqslant u\leqslant sT}\left|\int_u^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{\hat{s}\leqslant s-\delta\}}\right] \\ & +T^{-m/4}\mathbb{E}\left[\left|\sup_{sT\leqslant u\leqslant T}\int_{sT}^{u}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{\hat{s}\geqslant s+\delta\}}\right].\end{aligned}$$

By Cauchy-Schwartz inequality,

$$\begin{aligned}(A.92) \leqslant {} & T^{-m/4}\left\{\mathbb{E}\left[\sup_{0\leqslant u\leqslant sT}\left|\int_u^{sT}(\ln X(t))dB_t\right|^{m}\right]\mathrm{P}\{\hat{s}\leqslant s-\delta\}\right\}^{1/2} \\ & +T^{-m/4}\left\{\mathbb{E}\left[\sup_{sT\leqslant u\leqslant T}\left|\int_{sT}^{u}(\ln X(t))dB_t\right|^{m}\right]\mathrm{P}\{\hat{s}\geqslant s+\delta\}\right\}^{1/2}.\end{aligned}$$

By Burkholder-Davis-Gundy's inequality, there exists some constants $C_{m/2}$ such that

$$\begin{aligned}(A.92) \leqslant {} & T^{-m/4}\left\{C_{m/2}\mathbb{E}\left[\left(\int_{sT}^{T}(\ln X(t))^2dt\right)^{m/2}\right]\mathrm{P}\{\hat{s}\geqslant s+\delta\}\right\}^{1/2} \\ & +T^{-m/4}\left\{C_{m/2}\mathbb{E}\left[\left(\int_{0}^{sT}(\ln X(t))^2dt\right)^{m/2}\right]\mathrm{P}\{\hat{s}\leqslant s-\delta\}\right\}^{1/2}\end{aligned}$$

By Jensen inequality,

$$\left(\int_{sT}^{T}(\ln X(t))^2dt\right)^{m/2} \leqslant ((1-s)T)^{m/2-1}\int_{sT}^{T}|\ln X(t)|^m dt,$$

and then by Proposition 2.3.11,

$$\mathbb{E}\left[\left(\int_{sT}^{T}(\ln X(t))^2 dt\right)^{m/2}\right] \leqslant ((1-s)T)^{m/2-1}\int_{sT}^{T}\mathbb{E}\left[|\ln X(t)|^m\right]dt$$
$$\leqslant ((1-s)T)^{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right].$$

Similarly,

$$\mathbb{E}\left[\left(\int_{0}^{sT}(\ln X(t))^2 dt\right)^{m/2}\right] < (sT)^{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right].$$

This implies that

$$(A.92) \leqslant \frac{1}{T^{m/4}}\sqrt{C_{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\left(\sqrt{((1-s)T)^{m/2}}+\sqrt{(sT)^{m/2}}\right)\sqrt{\mathrm{P}\{|\hat{s}-s|>\delta\}}$$

$$= \sqrt{C_{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\left(\sqrt{(1-s)^{m/2}}+\sqrt{(s)^{m/2}}\right)\sqrt{\mathrm{P}\{|\hat{s}-s|>\delta\}}.$$

Then,

$$(A.92) \quad < 2\sqrt{C_{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}\sqrt{\varepsilon}. \tag{A.94}$$

Further,

$$(A.93) = \mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t - \frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{s-\delta<\hat{s}<s+\delta\}}\right]$$

$$= \mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_{\hat{s}T}^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{s-\delta<\hat{s}\leqslant s\}}\right] + \mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_{sT}^{\hat{s}T}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{s<\hat{s}\leqslant s+\delta\}}\right]$$

$$\leqslant T^{-m/4}\mathbb{E}\left[\sup_{(s-\delta)T<u<sT}\left|\int_u^{sT}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{s-\delta<\hat{s}\leqslant s\}}\right]$$
$$+T^{-m/4}\mathbb{E}\left[\sup_{sT<u<(s+\delta)T}\left|\int_{sT}^{u}(\ln X(t))dB_t\right|^{m/2}\mathbb{I}_{\{s<\hat{s}\leqslant s+\delta\}}\right]$$

By Cauchy-Schwartz inequality,

$$(A.93) \leqslant T^{-m/4}\left\{\mathbb{E}\left[\sup_{(s-\delta)T<t\leqslant sT}\left|\int_t^{sT}(\ln X(t))dB_t\right|^{m}\right]\mathrm{P}\{s-\delta<\hat{s}<s\}\right\}^{1/2}$$
$$+ T^{-m/4}\left\{\mathbb{E}\left[\sup_{sT<t\leqslant(s+\delta)T}\left|\int_{sT}^{t}(\ln X(t))dB_t\right|^{m}\right]\mathrm{P}\{s<\hat{s}\leqslant s+\delta\}\right\}^{1/2}$$
$$\leqslant T^{-m/4}\left\{\mathbb{E}\left[\sup_{(s-\delta)T<t\leqslant sT}\left|\int_t^{sT}(\ln X(t))dB_t\right|^{m}\right]\right\}^{1/2}$$

$$+ T^{-m/4}\left\{\mathbb{E}\left[\sup_{sT<t\leqslant(s+\delta)T}\left|\int_{sT}^{t}X(t)dB_t\right|^m\right]\right\}^{1/2}.$$

By Burkholder-Davis-Gundy's inequality, there exists some constants $C_{m/2}$ such that

$$\mathbb{E}\left[\sup_{(s-\delta)T<t\leqslant sT}\left|\int_{t}^{sT}(\ln X(t))dB_t\right|^m\right]\leqslant C_{m/2}\mathbb{E}\left[\left(\int_{(s-\delta)T}^{sT}(\ln X(t))^2dt\right)^{m/2}\right].$$

Further, from Jensen's inequality,

$$\left(\int_{(s-\delta)T}^{sT}(\ln X(t))^2dt\right)^{m/2}\leqslant(\delta T)^{m/2-1}\int_{(s-\delta)T}^{sT}|\ln X(t)|^m\,dt.$$

Then,

$$\mathbb{E}\left[\sup_{(s-\delta)T<t\leqslant sT}\left|\int_{t}^{sT}(\ln X(t))dB_t\right|^m\right]\leqslant C_{m/2}(\delta T)^{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right].$$

Similarly,

$$\mathbb{E}\left[\sup_{sT<t\leqslant(s+\delta)}\left|\int_{t}^{sT}(\ln X(t))dB_t\right|^m\right]\leqslant C_{m/2}(\delta T)^{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right].$$

Finally,

$$(A.93)\leqslant T^{-m/4}2\sqrt{C_{m/2}(\delta T)^{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}=2\sqrt{C_{m/2}\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^m\right]}(\delta)^{m/4}. \tag{A.95}$$

Since δ and ε can be arbitrary small, then, (A.94) and (A.95) imply that

$$\mathbb{E}\left[\left|\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t-\frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\right|^{m/2}\right]\xrightarrow[T\to\infty]{}0$$

which is

$$\frac{1}{\sqrt{T}}\int_0^{\hat{s}T}(\ln X(t))dB_t-\frac{1}{\sqrt{T}}\int_0^{sT}(\ln X(t))dB_t\xrightarrow[T\to\infty]{L^{m/2}}0.$$

This completes the proof. □

The following two propositions show that $\frac{1}{T}Q(\hat{s},T)-\frac{1}{T}Q(s,T)\xrightarrow[T\to\infty]{L^{m/2}}0.$

Proposition A.18. *Suppose that Assumption 2.1-2.3, and Assumption 2.4 hold and $\hat{s}$ is a consistent estimator of s, with $0<\hat{s}<1$ a.s. Then,*

(1). $\frac{1}{T}\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt-\frac{1}{T}\int_0^{sT}\varphi^\top(t)\varphi(t)dt\xrightarrow[T\to\infty]{L^m}\vec{0},$

(2) $\frac{1}{T}\int_0^{\hat{s}T}(\varphi(t), -\ln X(t))^\top(\ln X(t))dt - \frac{1}{T}\int_0^{sT}(\varphi(t), -\ln X(t))^\top(\ln X(t))dt \xrightarrow[T\to\infty]{L^{m/2}} \vec{0}.$

Proof. **(1)** The convergence will be proven in the following norms

$$\lim_{T\to\infty} T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m\right] = 0. \tag{A.96}$$

For some $K_\varphi > 0$, set$\|\varphi^\top(t)\varphi(t)\| \leqslant K_\varphi^2$ for all t. Let $\epsilon > 0$ and let $0 < \delta < \min\{s, 1-s\}$. For large enough T,

$$\mathrm{P}\left(|\hat{s} - s| \geqslant \frac{\delta}{K_\varphi^2}\right) < \varepsilon. \tag{A.97}$$

Let $G(T) = T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m\right]$, further, let

$$\begin{aligned} G_{11}(T) &= T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{\hat{s}>s\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\frac{\delta}{K_\varphi^2}\}}\right] \\ G_{12}(T) &= T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{\hat{s}>s\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\frac{\delta}{K_\varphi^2}\}}\right] \\ G_{21}(T) &= T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{\hat{s}<s\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\frac{\delta}{K_\varphi^2}\}}\right] \\ G_{22}(T) &= T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{\hat{s}<s\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\frac{\delta}{K_\varphi^2}\}}\right]. \end{aligned} \tag{A.98}$$

Note that (A.96) is equivalent to $\lim_{T\to\infty}(G_{11}(T) + G_{12}(T) + G_{21}(T) + G_{22}(T)) = 0$.

For the convergence of $G_{11}(T)$,

$$\begin{aligned} G_{11}(T) &= T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{\hat{s}>s\}}\mathbb{I}_{\{|\hat{s}-s|\leqslant\frac{\delta}{K_\varphi^2}\}}\right] \\ &= T^{-m}\mathbb{E}\left[\left\|\int_{sT}^{\hat{s}T}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{0<(\hat{s}-s)T\leqslant\frac{\delta T}{K_\varphi^2}\}}\right]. \end{aligned}$$

By Cauchy Schwartz's inequality,

$$\begin{aligned} G_{11}(T) &= T^{-m}\mathbb{E}\left[\left\|\int_{sT}^{\hat{s}T}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{0<(\hat{s}-s)T\leqslant\frac{\delta T}{K_\varphi^2}\}}\right] \\ &\leqslant T^{-m}\mathbb{E}\left[\left|\int_{sT}^{\hat{s}T}\left\|\varphi^\top(t)\varphi(t)\right\|dt\right|^m \mathbb{I}_{\{0<(\hat{s}-s)T\leqslant\frac{\delta T}{K_\varphi^2}\}}\right] \\ &\leqslant T^{-m}\mathbb{E}\left[(K_\varphi^2(\hat{s}-s)T)^m\mathbb{I}_{\{0<(\hat{s}-s)T\leqslant\frac{\delta T}{K_\varphi^2}\}}\right] \leqslant \delta^m \end{aligned}$$

which implies that $\lim_{T\to\infty} G_{11}(T) = 0$. Further, as $0 < s < 1$ and $0 < \hat{s} < 1$ a.s.,

$$G_{12}(T) = T^{-m}\mathbb{E}\left[\left\|\int_0^{\hat{s}T}\varphi^\top(t)\varphi(t)dt - \int_0^{sT}\varphi^\top(t)\varphi(t)dt\right\|^m \mathbb{I}_{\{1>\hat{s}-s>0\}}\mathbb{I}_{\{|\hat{s}-s|\geqslant\frac{\delta}{4pK_\varphi}\}}\right]$$

$$\leqslant T^{-m}\mathbb{E}\left[\left\|\int_{sT}^{\hat{s}T}\varphi^{\top}(t)\varphi(t)dt\right\|^{m}\mathbb{I}_{\{T>(\hat{s}-s)T>0\}}\mathbb{I}_{\{\hat{s}-s\geqslant\frac{\delta}{K_{\varphi}^{2}}\}}\right]$$

$$\leqslant T^{-m}\mathbb{E}\left[(K_{\varphi}^{2}|\hat{s}-s|T)^{m}\mathbb{I}_{\{T>(\hat{s}-s)T\geqslant\frac{\delta T}{K_{\varphi}^{2}}\}}\right]=K_{\varphi}^{2m}\mathbb{E}\left[|\hat{s}-s|^{m}\mathbb{I}_{\{T>(\hat{s}-s)T\geqslant\frac{\delta T}{K_{\varphi}^{2}}\}}\right]\xrightarrow[T\to\infty]{}0.$$

Furthermore, since $|\hat{s}-s|\leqslant 2$ almost surely,

$$G_{12}(T)\leqslant K_{\varphi}^{2m}2^{m}\mathbb{E}\left[\mathbb{I}_{\{T>(\hat{s}-s)T\geqslant\frac{\delta T}{K_{\varphi}^{2}}\}}\right]=K_{\varphi}^{2m}2^{m}\mathrm{P}\left\{\hat{s}-s\geqslant\frac{\delta}{K_{\varphi}^{2}}\right\}\xrightarrow[T\to\infty]{}0.$$

The using of similar techniques as for $G_{11}(T)$ and $G_{12}(T)$ implies that $\lim_{T\to\infty}G_{21}(T)=\lim_{T\to\infty}G_{22}(T)=0$. This proves that (A.96) hold.

(2) Similarly to the proof of **(1)**, one can prove that

$$\frac{1}{T}\int_{0}^{\hat{s}T}(\varphi(t),-\ln X(t))^{\top}(\ln X(t))dt-\frac{1}{T}\int_{0}^{sT}(\varphi(t),-\ln X(t))^{\top}(\ln X(t))dt\xrightarrow[T\to\infty]{L^{m/2}}\vec{0}_{(p_0+1)\times 1}$$

under the fact that $\sup_{t\geqslant 0}\mathbb{E}\left[|\ln X(t)|^{m}\right]<\infty$, for $m\geqslant 2$. This completes the proof.

□

Proof of Proposition 2.4.1. By Assumption 2.3 the basis function $\varphi(t)$ is Riemann integrable in the interval $[0,T]$, which implies that $\varphi(t),t\in[0,T]$ is continuous almost everywhere. Since the function $\varphi^{\top}(t)\varphi(t)$ is an almost everywhere continuous function, by Proposition A.13 and Assumption 2.4,

$$\begin{aligned}&\mathbb{E}\left[\left\|\sum_{k=0}^{N-1}\varphi(t_k)\varphi^{\top}(t_k)(t_{k+1}-t_k)-\int_{0}^{T}\varphi(t)\varphi^{\top}(t)dt\right\|^{m}\right]\\&\leqslant N^{m-1}\sum_{k=0}^{N-1}\left(\int_{t_k}^{t_{k+1}}\left\|\left(\varphi_i(t_k)\varphi_j(t_k)-\varphi_i(t)\varphi_j(t)\right)\right\|dt\right)^{m}\leqslant(C_1(\Delta_N))^{m}(N\Delta_N)^{m},\end{aligned}\tag{A.99}$$

where $C_1(\Delta_N)$ is defined in Proposition A.13. Further,

$$\begin{aligned}&\mathbb{E}\left[\left\|\sum_{k=0}^{N-1}\varphi(t_k)\ln X(t_k)\Delta_N-\int_{0}^{T}\varphi(t)\ln X(t)dt\right\|^{m}\right]\\&\leqslant N^{m-1}2^{m-1}\sum_{k=0}^{N-1}\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\varphi(t_k)(\ln X(t_k)-\ln X(t))dt\right\|^{m}\right]\\&+N^{m-1}2^{m-1}\sum_{k=0}^{N-1}\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\ln X(t)(\varphi(t_k)-\varphi(t))dt\right\|^{m}\right].\end{aligned}$$

Further, the upper bound of $\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\varphi(t_k)(\ln X(t_k)-\ln X(t))dt\right\|^m\right]$ and $\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\ln X(t)(\varphi(t_k)-\varphi(t))dt\right\|^m\right]$ need to be found. Since $m \geqslant 2$, by Jensen Inequality,

$$\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\varphi(t_k)(\ln X(t_k)-\ln X(t))dt\right\|^m\right] \leqslant (t_{k+1}-t_k)^{m-1}K_\varphi^m\int_{t_k}^{t_{k+1}}\mathbb{E}\left[|\ln X(t_k)-\ln X(t)|^m\right]dt.$$

By Proposition 2.2.1,

$$|\ln X(t_k)-\ln X(t)|^m \leqslant 3^{m-1}\left(|e^{-\alpha t_k}-e^{-\alpha t}|^m|\ln X_0|^m+|r(t_k)-r(t)|^m+|\tau(t_k)-\tau(t)|^m\right).$$

Since the function e^x is differentiable at $x\in\mathbb{R}$, by the Taylor series expansion,

$$|e^{-\alpha t}-e^{-\alpha t_k}| \leqslant e^{-\alpha t_k}\alpha|t-t_k|+\frac{e^{-\alpha\xi_1}}{2!}(\alpha)^2(t-t_k)^2,$$

for some ξ_1 between t and t_k. This implies that

$$\begin{aligned}&\mathbb{E}\left[(|e^{-\alpha t_k}-e^{-\alpha t}||\ln X_0|)^m\right] \leqslant \left(e^{-\alpha t_k}\alpha|t-t_k|+\frac{e^{-\alpha\xi_1}}{2!}(\alpha)^2(t-t_k)^2\right)^m\mathbb{E}\left[|\ln X_0|^m\right]\\&\leqslant\left(\alpha\Delta_N+\frac{\alpha^2}{2!}\Delta_N^2\right)^m\mathbb{E}\left[|\ln X_0|^m\right]=\mathbb{E}\left[|\ln X_0|^m\right]2^{m-1}\left((\alpha\Delta_N)^m+\left(\frac{\alpha^2}{2!}\Delta_N^2\right)^m\right).\end{aligned} \tag{A.100}$$

Further, for $t_k \leqslant t \leqslant t_{k+1}$,

$$\begin{aligned}|r(t_k)-r(t)| &= \left|\sum_{j=1}^p\mu_j\int_0^{t_k}e^{-\alpha(t_k-s)}\varphi_j(s)ds-\sum_{j=1}^p\mu_j\int_0^t e^{-\alpha(t-s)}\varphi_j(s)ds\right|\\&=\left|\sum_{j=1}^p\mu_j\int_0^{t_k}(e^{-\alpha(t_k-s)}-e^{-\alpha(t-s)})\varphi_j(s)ds-\sum_{j=1}^p\mu_j\int_{t_k}^t e^{-\alpha(t-s)}\varphi_j(s)ds\right|\\&\leqslant\left|\sum_{j=1}^p\mu_j\int_0^{t_k}(e^{-\alpha(t_k-s)}-e^{-\alpha(t-s)})\varphi_j(s)ds\right|+\left|\sum_{j=1}^p\mu_j\int_{t_k}^t e^{-\alpha(t-s)}\varphi_j(s)ds\right|.\end{aligned}$$

Then,

$$|r(t_k)-r(t)| \leqslant \sum_{j=1}^p|\mu_j|\int_0^{t_k}(e^{-\alpha(t_k-s)}-e^{-\alpha(t-s)})\left|\varphi_j(s)\right|ds+\sum_{j=1}^p|\mu_j|\int_{t_k}^t e^{-\alpha(t-s)}\left|\varphi_j(s)\right|ds,\ t\geqslant t_k.$$

From the fact $\sum_{j=1}^p|\mu_j| \leqslant K_\mu < \infty$ and $|\varphi_j(t)| \leqslant K_\varphi < \infty$ for $k=1,2,\cdots,p, t\geqslant 0$,

$$\begin{aligned}|r(t_k)-r(t)| &\leqslant K_\mu K_\varphi\left[\left|\int_0^{t_k}(e^{-\alpha(t_k-s)}-e^{-\alpha(t-s)})ds\right|+\left|\int_{t_k}^t e^{-\alpha(t-s)}ds\right|\right]\\&=K_\mu K_\varphi\left[\frac{1}{\alpha}\left|1-e^{-\alpha t_k}-e^{-\alpha(t-t_k)}+e^{-\alpha t}\right|+\frac{1}{\alpha}\left|1-e^{-\alpha(t-t_k)}\right|\right],\quad t\geqslant t_k\end{aligned}$$

$$= \frac{K_\mu K_\varphi}{\alpha}\left[\left|1 - e^{-\alpha(t-t_k)} + e^{-\alpha t} - e^{-\alpha t_k}\right| + \left|1 - e^{-\alpha(t-t_k)}\right|\right], \quad t \geqslant t_k.$$

This gives

$$|r(t_k) - r(t)| \leqslant \frac{K_\mu K_\varphi}{\alpha}\left[\left|e^{-\alpha t} - e^{-\alpha t_k}\right| + 2\left|1 - e^{-\alpha(t-t_k)}\right|\right], \quad t \geqslant t_k.$$

By Taylor series expansion, for $t \geqslant t_k$,

$$e^{-\alpha t} - e^{-\alpha t_k} = e^{-\alpha t_k}(-\alpha)(t - t_k) + \frac{e^{-\alpha\xi_3}}{2!}(\alpha)^2(t - t_k)^2,$$

$$e^{-\alpha(t-t_k)} - 1 = e^{-\alpha t_k}(-\alpha)(t - t_k) + \frac{e^{-\alpha\xi_2}}{2!}(\alpha)^2(t - t_k)^2,$$

for some ξ_2, ξ_3 in the interval $[t_k, t]$. By using $e^{-\alpha\xi_2} \leqslant e^{-\alpha t_k}$ and $e^{-\alpha\xi_3} \leqslant e^{-\alpha t_k}$, then

$$\begin{aligned} |r(t_k) - r(t)|^m &\leqslant \left(\frac{K_\mu K_\varphi}{\alpha}\right)^m \left(3\alpha e^{-\alpha t_k}|t - t_k| + \tfrac{3}{2}e^{-\alpha t_k}\alpha^2(t - t_k)^2\right)^m \\ &\leqslant \left(\frac{K_\mu K_\varphi}{\alpha}\right)^m \left(3\alpha\Delta_N + \tfrac{3}{2}\alpha^2\Delta_N^2\right)^m = \left(\frac{K_\mu K_\varphi}{\alpha}\right)^m 2^{m-1}\left((3\alpha\Delta_N)^m + \left(\tfrac{3}{2}\alpha^2\Delta_N^2\right)^m\right), \end{aligned} \tag{A.101}$$

since $t - t_k < t_{k+1} - t_k$ and $e^{-2\alpha t_k} < 1$. Then, for the last term, by Cauchy-Schwartz inequality, $\mathbb{E}[|\tau(t_k) - \tau(t)|] \leqslant \{\mathbb{E}[|\tau(t_k) - \tau(t)|^2]\}^{1/2}$ and since $t > t_k$,

$$\mathbb{E}[|\tau(t_k) - \tau(t)|^m] = \mathbb{E}\left[\left|\sigma e^{-\alpha t_k}\int_0^{t_k} e^{\alpha s} dB_s - \sigma e^{-\alpha t}\int_0^t e^{\alpha s} dB_s\right|^m\right]$$

$$= \sigma^m \mathbb{E}\left[\left|(e^{-\alpha t_k} - e^{-\alpha t})\int_0^{t_k} e^{\alpha s} dB_s - e^{-\alpha t}\int_{t_k}^t e^{\alpha s} dB_s\right|^m\right]$$

$$\leqslant \sigma^m 2^{m-1}\left((e^{-\alpha t_k} - e^{-\alpha t})^m\, \mathbb{E}\left[\left|\int_0^{t_k} e^{\alpha s} dB_s\right|^m\right] + e^{-\alpha m t}\mathbb{E}\left[\left|\int_{t_k}^t e^{\alpha s} dB_s\right|^m\right]\right),$$

for $t \geqslant t_k$. Further, by Burkholder-Davis-Gundy's inequality, there exists some positive constant C_m, such that

$$\mathbb{E}\left[\left|\int_0^{t_k} e^{\alpha s} dB_s\right|^m\right] \leqslant C_m \mathbb{E}\left[\left(\int_0^{t_k} e^{2\alpha s} ds\right)^{m/2}\right] = C_m\left(\frac{1}{2\alpha}(e^{2\alpha t_k} - 1)\right)^{m/2},$$

and

$$\mathbb{E}\left[\left|\int_{t_k}^t e^{\alpha s} dB_s\right|^m\right] \leqslant C_m \mathbb{E}\left[\left(\int_{t_k}^t e^{2\alpha s} ds\right)^{m/2}\right] = C_m\left(\frac{1}{2\alpha}(e^{2\alpha t} - e^{2\alpha t_k})\right)^{m/2}.$$

Then, $(e^{-\alpha t_k} - e^{-\alpha t})^m \mathbb{E}\left[\left|\int_0^{t_k} e^{\alpha s} dB_s\right|^m\right] \leqslant C_m\left(\frac{1}{2\alpha}(e^{-\alpha t_k} - e^{-\alpha t})^2(e^{2\alpha t_k} - 1)\right)^{m/2}$. By Taylor series expansion again,

$$
\begin{aligned}
(e^{-\alpha t_k}-e^{-\alpha t})^2 &= \left(e^{-\alpha t_k}(-\alpha)(t-t_k)+\frac{e^{-\alpha\xi_3}}{2!}(\alpha)^2(t-t_k)^2\right)^2\\
&\leqslant 2\left(e^{-2\alpha t_k}\alpha^2(t-t_k)^2+e^{-2\alpha\xi_3}(\alpha)^4(t-t_k)^4\right)\\
&\leqslant 2\left(e^{-2\alpha t_k}\alpha^2(t-t_k)^2+e^{-2\alpha t_k}(\alpha)^4(t-t_k)^4\right),
\end{aligned}
$$

which implies that

$$
\begin{aligned}
(e^{-\alpha t_k}-e^{-\alpha t})^2\frac{1}{2\alpha}(e^{2\alpha t_k}-1) &\leqslant \frac{1}{2\alpha}2\left(e^{-2\alpha t_k}\alpha^2(t-t_k)^2+e^{-2\alpha t_k}(\alpha)^4(t-t_k)^4\right)\\
&+\frac{1}{2\alpha}(e^{2\alpha t_k})2\left(e^{-2\alpha t_k}\alpha^2(t-t_k)^2+e^{-2\alpha t_k}(\alpha)^4(t-t_k)^4\right)\\
&= \frac{1}{\alpha}\left(e^{-2\alpha t_k}\alpha^2(t-t_k)^2+e^{-2\alpha t_k}(\alpha)^4(t-t_k)^4+\alpha^2(t-t_k)^2+(\alpha)^4(t-t_k)^4\right)\\
&= \frac{1}{\alpha}\left((e^{-2\alpha t_k}+1)\alpha^2(t-t_k)^2+(e^{-2\alpha t_k}+1)(\alpha)^4(t-t_k)^4\right).\\
&\leqslant \frac{1}{\alpha}\left(2\alpha^2(t-t_k)^2+2(\alpha)^4(t-t_k)^4\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&(e^{-\alpha t_k}-e^{-\alpha t})^m\mathbb{E}\left[\left(\int_0^{t_k}e^{\alpha s}dB_s\right)^m\right]\leqslant C_m\left(\frac{1}{\alpha}\left(2\alpha^2(t-t_k)^2+2(\alpha)^4(t-t_k)^4\right)\right)^{m/2}\\
&\leqslant C_m\left(\frac{1}{\alpha}\left(2\alpha^2(\Delta_N)^2+2(\alpha)^4(\Delta_N)^4\right)\right)^{m/2}.
\end{aligned}
\tag{A.102}
$$

Further,

$$
\begin{aligned}
e^{-\alpha m t}\mathbb{E}\left[\left|\int_{t_k}^{t}e^{\alpha s}dB_s\right|^m\right] &\leqslant C_m e^{-\alpha m t}\mathbb{E}\left[\left(\int_{t_k}^{t}e^{2\alpha s}ds\right)^{m/2}\right]\\
= C_m\left(\frac{1}{2\alpha}e^{-2\alpha t}(e^{2\alpha t}-e^{2\alpha t_k})\right)^{m/2} &= C_m\left(\frac{1}{2\alpha}(1-e^{-2\alpha(t-t_k)})\right)^{m/2},
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{1}{2\alpha}\left(1-e^{-2\alpha(t-t_k)}\right) &\leqslant \frac{1}{2\alpha}\left(e^{-2\alpha t_k}(2\alpha)(t-t_k)+\frac{e^{-2\alpha t_k}}{2!}(2\alpha)^2(t-t_k)^2\right)\\
\leqslant \frac{1}{\alpha}\left(e^{-2\alpha t_k}(\alpha)(t-t_k)+\frac{e^{-2\alpha t_k}}{2!}2(\alpha)^2(t-t_k)^2\right) &\leqslant \alpha\Delta_N+(\alpha)^2(\Delta_N)^2,\quad t\geqslant t_k.
\end{aligned}
$$

Then,

$$C_m\left(\frac{1}{2\alpha}(1-e^{-2\alpha(t-t_k)})\right)^{m/2} \leqslant C_m\left(\alpha\Delta_N+(\alpha)^2(\Delta_N)^2\right)^{m/2}. \tag{A.103}$$

(A.102) and (A.103) imply that

$$\mathbb{E}[|\tau(t_k)-\tau(t)|^m] \leqslant \sigma^m 2^{m-1}C_m\left(\left(\frac{1}{\alpha}\left(2\alpha^2(\Delta_N)^2+2(\alpha)^4(\Delta_N)^4\right)\right)^{m/2}+\left(\alpha\Delta_N+(\alpha)^2(\Delta_N)^2\right)^{m/2}\right). \tag{A.104}$$

Since $t-t_k<t_{k+1}-t_k$ and $e^{-2\alpha t_k}<1$, by (A.100), (A.101) and (A.104),

$$\begin{aligned}\mathbb{E}\left[|\ln X(t_k)-\ln X(t)|^m\right] &\leqslant 3^{m-1}\left(\mathbb{E}\left[|\ln X_0|^m\right]2^{m-1}\left((\alpha\Delta_N)^m+\left(\frac{\alpha^2}{2!}\Delta_N^2\right)^m\right)\right.\\ &\quad+\left(\frac{K_\mu K_\varphi}{\alpha}\right)^m 2^{m-1}\left((3\alpha\Delta_N)^m+\left(\tfrac{3}{2}\alpha^2\Delta_N^2\right)^m\right)\\ &\quad+\sigma^m 2^{m-1}C_m\left(\left(\frac{1}{\alpha}\left(2\alpha^2(\Delta_N)^2+2(\alpha)^4(\Delta_N)^4\right)\right)^{m/2}\right.\\ &\quad\left.\left.+\left(\alpha\Delta_N+(\alpha)^2(\Delta_N)^2\right)^{m/2}\right)\right).\end{aligned} \tag{A.105}$$

Then,

$$\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\varphi(t_k)(\ln X(t_k)-\ln X(t))dt\right\|^m\right] \leqslant (t_{k+1}-t_k)^{m-1}K_\varphi^m\int_{t_k}^{t_{k+1}}\mathbb{E}\left[|\ln X(t_k)-\ln X(t)|^m\right]dt$$

$$\leqslant (t_{k+1}-t_k)^{m-1}K_\varphi^m 3^{m-1}\left\{\left(\mathbb{E}\left[|\ln X_0|^m\right](\alpha)^m+\left(K_\mu K_\varphi\right)^m 3^m\right)(\Delta_N)^m+\sigma^m 2^{m-1}C_m\left((2\alpha)^m(\Delta_N)^m\right.\right.$$

$$\left.\left.+(\alpha)^{m/2}(\Delta_N)^{m/2}\right)+o\left((\Delta_N)^m\right)\right\}(t_{k+1}-t_k)$$

$$\leqslant (\Delta_N)^m K_\varphi^m 3^{m-1}\left\{\left(\mathbb{E}\left[|\ln X_0|^m\right](\alpha)^m+\left(K_\mu K_\varphi\right)^m 3^m\right)(\Delta_N)^m+\sigma^m 2^{m-1}C_m\left((2\alpha)^m(\Delta_N)^m\right.\right.$$

$$\left.\left.+(\alpha)^{m/2}(\Delta_N)^{m/2}\right)+o\left((\Delta_N)^m\right)\right\}=(\Delta_N)^m K_\varphi^m 3^{m-1}\left(f(m,\Delta_N)+o\left((\Delta_N)^m\right)\right),$$

where

$$\begin{aligned}f(m,\Delta_N) &= \left(\mathbb{E}\left[|\ln X_0|^m\right](\alpha)^m+\left(K_\mu K_\varphi\right)^m 3^m\right)(\Delta_N)^m\\ &\quad+\sigma^m 2^{m-1}C_m\left((2\alpha)^m(\Delta_N)^m+(\alpha)^{m/2}(\Delta_N)^{m/2}\right).\end{aligned}$$

By Jensen inequality, Proposition 2.2.2 and Proposition A.13,

$$\begin{aligned}\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}\ln X(t)(\varphi_i(t_k)-\varphi_i(t))dt\right\|^m\right] &\leqslant (t_{k+1}-t_k)^{m-1}\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]\times\\ &\quad\int_{t_k}^{t_{k+1}}\|\varphi(t_k)-\varphi(t)\|^m\,dt\end{aligned}$$

$$\leqslant \sup_{t\geqslant 0} \mathbb{E}[|\ln X(t)|^m]\,(C_2(\Delta_N))^m\,(\Delta_N)^m.$$

So,

$$\mathbb{E}\left[\left\|\sum_{k=0}^{N-1} \varphi_i(t_k)\ln X(t_k)\Delta_N - \int_0^T \varphi_i(t)\ln X(t)dt\right\|^m\right]$$
$$\leqslant (N\Delta_N)^m 2^{m-1}\left(K_\varphi^m 3^{m-1}\left(f(m,\Delta_N)+o\left((\Delta_N)^m\right)\right)+\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]\,(C_2(\Delta_N))^m\right). \tag{A.106}$$

For the term $\ln X(t)$, by Jensen's inequality and triangle inequality,

$$\mathbb{E}\left[\left|\sum_{k=0}^{N-1}(\ln X(t_k))^2\Delta_N - \int_0^T(\ln X(t))^2dt\right|^{m/2}\right]$$
$$\leqslant N^{m/2-1}\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}}\left|\left(\ln X(t_k))^2-(\ln X(t))^2\right)\right|dt\right)^{m/2}\right]$$
$$\leqslant N^{m/2-1}2^{m/2-1}\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}}|\ln X(t_k)|\,|\ln X(t_k)-\ln X(t)|\,dt\right)^{m/2}\right]$$
$$+N^{m-1}2^{m-1}\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}}|\ln X(t)|\,|\ln X(t_k)-\ln X(t)|\,dt\right)^{m/2}\right]$$
$$\leqslant N^{m/2-1}2^{m/2-1}\sum_{k=0}^{N-1}(t_{k+1}-t_k)^{m-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\left[(|\ln X(t_k)|\,|\ln X(t_k)-\ln X(t)|)^{m/2}\right]dt$$
$$+N^{m/2-1}2^{m/2-1}\sum_{k=0}^{N-1}(t_{k+1}-t_k)^{m-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\left[(|\ln X(t)|\,|\ln X(t_k)-\ln X(t)|)^{m/2}\right]dt.$$

Then, by Cauchy-Schwartz inequality,

$$\mathbb{E}\left[(|\ln X(t_k)|\,|\ln X(t_k)-\ln X(t)|)^{m/2}\right]\leqslant\left\{\mathbb{E}\left[|\ln X(t_k)|^m\right]\mathbb{E}\left[|\ln X(t_k)-\ln X(t)|^m\right]\right\}^{1/2},$$

$$\mathbb{E}\left[(|\ln X(t)|\,|\ln X(t_k)-\ln X(t)|)^{m/2}\right]\leqslant\left\{\mathbb{E}\left[|\ln X(t)|^m\right]\mathbb{E}\left[|\ln X(t_k)-\ln X(t)|^m\right]\right\}^{1/2}.$$

By (A.105) and (A.55),

$$
\begin{aligned}
&\mathbb{E}\left[\left|\sum_{k=0}^{N-1}(\ln X(t_k))^2\Delta_N - \int_0^T (\ln X(t))^2\,dt\right|^{m/2}\right]\\
&\leqslant N^{m/2-1}2^{m/2-1}\sqrt{\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]}\sum_{k=0}^{N-1}(t_{k+1}-t_k)^{m/2-1}\int_{t_k}^{k+1}\sqrt{3^{m-1}(f(m,\Delta_N)+o\left((\Delta_N)^m\right))}dt\\
&\leqslant N^{m/2-1}2^{m/2-1}\sqrt{\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]}\sqrt{3^{m-1}(f(m,\Delta_N)+o\left((\Delta_N)^m\right))}N(\Delta_N)^{m/2}\\
&\leqslant (N\Delta_N)^{m/2}2^{m/2-1}\sqrt{\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]}\sqrt{3^{m-1}(f(m,\Delta_N)+o\left((\Delta_N)^m\right))}.
\end{aligned}
\tag{A.107}
$$

(A.99), (A.106) and (A.107) imply that

$$
\begin{aligned}
&\left\|\sum_{k=0}^{N-1}\Psi(t_k)\Psi^\top(t_k)(t_{k+1}-t_k) - \int_0^T \Psi(t)\Psi^\top(t)dt\right\|_1^{m/2} \leqslant \max\Big\{(C_1(\Delta_N))^{m/2}(N\Delta_N)^{m/2},\\
&(N\Delta_N)^{m/2}\sqrt{2^{m-1}\left(K_\varphi^m 3^{m-1}\left(f(m,\Delta_N)+o\left((\Delta_N)^m\right)\right)+\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]\left(C_2(\Delta_N)\right)^m\right)},\\
&(N\Delta_N)^{m/2}2^{m/2-1}\sqrt{\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]}\sqrt{3^{m-1}(f(m,\Delta_N)+o\left((\Delta_N)^m\right))}\Big\}.
\end{aligned}
\tag{A.108}
$$

(A.108) still holds if 1-*norm* is replaced by ∞-*norm*. Further, since for a matrix A, $\|A\|_2 \leqslant \|A\|_1\|A\|_\infty$, by Assumption 2.4,

$$
\left\|\sum_{k=0}^{N-1}\Psi(t_k)\Psi^\top(t_k)\Delta_N - \int_0^T \Psi(t)\Psi^\top(t)dt\right\|_2 \leqslant K(m,\Delta_N)O(T^{m/2}),
$$

where

$$
\begin{aligned}
K(m,\Delta_N) = \max\Big\{&(C_1(\Delta_N))^{m/2},\\
&\sqrt{2^{m-1}\left(K_\varphi^m 3^{m-1}\left(f(m,\Delta_N)+o\left((\Delta_N)^m\right)\right)+\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]\left(C_2(\Delta_N)\right)^m\right)},\\
&2^{m/2-1}\sqrt{\sup_{t\geqslant 0}\mathbb{E}[|\ln X(t)|^m]}\sqrt{3^{m-1}(f(m,\Delta_N)+o\left((\Delta_N)^m\right))}\Big\}.
\end{aligned}
$$

Further, $\|A\|_F \leqslant \sqrt{p+1}\|A\|_2$, where $p+1$ is the rank of matrix A. This proves assertion (*i*).

For the assertion (*ii*), by Minkowsky's inequality,

$$\mathbb{E}\left[\left|\sum_{k=0}^{N-1}\ln X(t_k)(B_{t_{k+1}}-B_{t_k})-\int_0^T \ln X(t)dB_t\right|^m\right]=\mathbb{E}\left[\left|\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}(\ln X(t_k)-\ln X(s))dB_s\right|^m\right]$$

$$\leqslant \sum_{k=0}^{N-1}\mathbb{E}\left[\left|\int_{t_k}^{t_{k+1}}(\ln X(t_k)-\ln X(s))dB_s\right|^m\right],$$

and by Burkholder-Davis-Gundy and Jensen inequalities, for some positive constant C_m,

$$\sum_{k=0}^{N-1}\mathbb{E}\left[\left|\int_{t_k}^{t_{k+1}}(\ln X(t_k)-\ln X(s))dB_s\right|^m\right]\leqslant C_m\sum_{k=0}^{N-1}\mathbb{E}\left[\left(\int_{t_k}^{t_{k+1}}(\ln X(t_k)-\ln X(s))^2ds\right)^{m/2}\right]$$

$$\leqslant C_m\sum_{k=0}^{N-1}(t_{k+1}-t_k)^{m/2-1}\int_{t_k}^{t_{k+1}}\mathbb{E}\left[|\ln X(t_k)-\ln X(s)|^m\right]ds.$$

Then, together with (A.105),

$$\mathbb{E}\left[|\ln X(t_k)-\ln X(t)|^m\right]\leqslant 3^{m-1}\Bigg(\mathbb{E}\left[|\ln X_0|^m\right]2^{m-1}\left((\alpha\Delta_N)^m+\left(\frac{\alpha^2}{2!}\Delta_N^2\right)^m\right)$$
$$+\left(\frac{K_\mu K_\varphi}{\alpha}\right)^m 2^{m-1}\left((3\alpha\Delta_N)^m+\left(\tfrac{3}{2}\alpha^2\Delta_N^2\right)^m\right)+\sigma^m 2^{m-1}C_m\Bigg(\left(\frac{1}{\alpha}\left(2\alpha^2(\Delta_N)^2+2(\alpha)^4(\Delta_N)^4\right)\right)^{m/2}$$
$$+\left(\alpha\Delta_N+(\alpha)^2(\Delta_N)^2\right)^{m/2}\Bigg)\Bigg).$$

Then,

$$\sum_{k=0}^{N-1}\mathbb{E}\left[\left|\int_{t_k}^{t_{k+1}}(\ln X(t_k)-\ln X(s))dB_s\right|^m\right]\leqslant C_m(\Delta_N)^{m/2-1}3^{m-1}\Bigg(\mathbb{E}\left[|\ln X_0|^m\right]2^{m-1}\Big((\alpha\Delta_N)^m$$
$$+\left(\frac{\alpha^2}{2!}\Delta_N^2\right)^m\Big)+\left(\frac{K_\mu K_\varphi}{\alpha}\right)^m 2^{m-1}\left((3\alpha\Delta_N)^m+\left(\tfrac{3}{2}\alpha^2\Delta_N^2\right)^m\right)+\sigma^m 2^{m-1}C_m\Bigg(\Big(\frac{1}{\alpha}\Big(2\alpha^2(\Delta_N)^2$$
$$+2(\alpha)^4(\Delta_N)^4\Big)\Big)^{m/2}+\left(\alpha\Delta_N+(\alpha)^2(\Delta_N)^2\right)^{m/2}\Bigg)\Bigg)T$$
$$=C_m(\Delta_N)^{m/2-1}3^{m-1}\left((\alpha\Delta_N)^{m/2}+o((\Delta_N)^{m/2})\right)T. \tag{A.109}$$

Further,

$$\mathbb{E}\left[\left\|\sum_{k=0}^{N-1}\varphi(t_k)(B_{t_{k+1}}-B_{t_k})-\int_0^T\varphi(t)dB_t\right\|^m\right]=\mathbb{E}\left[\left\|\sum_{k=0}^{N-1}\int_{t_k}^{t_{k+1}}(\varphi(t_k)-\varphi(s))dB_s\right\|^m\right]$$

$$\leqslant\sum_{k=0}^{N-1}\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}(\varphi(t_k)-\varphi(s))dB_s\right\|^m\right],$$

and by Burkholder-Davis-Gundy inequality, for some positive constant C_m,

$$\sum_{k=0}^{N-1}\mathbb{E}\left[\left\|\int_{t_k}^{t_{k+1}}(\varphi(t_k)-\varphi(s))dB_s\right\|^m\right]\leqslant\sum_{k=0}^{N-1}\mathbb{E}\left[\sup_{t_k\leqslant u\leqslant t_{k+1}}\left\|\int_{t_k}^{u}(\varphi_j(t_k)-\varphi_j(s))dB_s\right\|^m\right]$$

$$\leqslant C_m\sum_{k=0}^{N-1}\left\|\int_{t_k}^{t_{k+1}}(\varphi(t_k)-b(s))^\top(\varphi(t_k)-\varphi(s))ds\right\|^{m/2}.$$

By Assumption 2.3, the base function $\varphi(t)$ is Riemann integrable on the interval $[0,T]$, it is continuous almost everywhere. By Proposition A.13, there exists a nonnegative and non decreasing function $C_3(s-t_k)$, with $\inf_{\{s-t_k>0\}}C_3(s-t_k)=0$, such that $\|\varphi(t_k)-\varphi(s)\|\leqslant C_3(s-t_k)$.

$$C_m\sum_{k=0}^{N-1}\left\|\int_{t_k}^{t_{k+1}}(\varphi(t_k)-\varphi(s))^\top(\varphi(t_k)-\varphi(s))ds\right\|^{m/2}\leqslant C_m\left(C_3(\Delta_N)\right)^m(\Delta_N)^{m/2}\,T. \tag{A.110}$$

(A.109) and (A.110) complete the proof of assertion (*ii*). □

Proof of Proposition 2.4.3. Since SDE (2.2.1) admits a strong and unique solution that uniformly bounded in L^m. i.e. $\exists\ K>0$, such that $\sup_{t\geqslant0}\mathbb{E}[(\ln X(t))^2]\leqslant K$,

$$\int_0^T\mathbb{E}[(\ln X(t))^2]dt\leqslant KT. \tag{A.111}$$

Then, by combining Markov's inequality,

$$\mathrm{P}\left(\frac{1}{\sqrt{T}}\left|\int_0^T\ln X(t)dB_t\right|>K^*\right)\leqslant\frac{\mathbb{E}\left(\left|\int_0^T\ln X(t)dB_t\right|^2\right)}{T(K^*)^2},$$

by Itô's isometry,

$$\frac{\mathbb{E}\left(\left|\int_0^T\ln X(t)dB_t\right|^2\right)}{T(K^*)^2}=\frac{\int_0^T\mathbb{E}[(\ln X(t))^2]dt}{T(K^*)^2},$$

and by inequality of (A.111),

$$\frac{\int_0^T\mathbb{E}[(\ln X(t))^2]dt}{T(K^*)^2}\leqslant\frac{KT}{T(K^*)^2}=\frac{K}{(K^*)^2}.$$

$$\mathrm{P}\left(\frac{1}{\sqrt{T}}\left|\int_0^T \ln X(t)dB_t\right| > K^*\right) \leqslant \frac{\mathbb{E}\left(\left|\int_0^T \ln X(t)dB_t\right|^2\right)}{T(K^*)^2} = \frac{K}{(K^*)^2}.$$

Further, by combining Markov's inequality, Itô's isometry and Assumption 2.3

$$\mathrm{P}\left(\frac{1}{\sqrt{T}}\left\|\int_0^T \varphi(t)dB_t\right\| > K^*\right) \leqslant \frac{\mathbb{E}\left(\left\|\int_0^T \varphi(t)dB_t\right\|^2\right)}{T(K^*)^2} = \frac{\int_0^T \|\varphi(t)\|^2 dt}{T(K^*)^2} \leqslant \frac{K_\varphi T}{T(K^*)^2} = \frac{K_\varphi}{(K^*)^2}.$$

Set $K^* = \log^{a^*} T$. For some $0 < a^* < a/2$, that

$$\mathrm{P}\left(\frac{1}{\sqrt{T}}\left\|\int_0^T \varphi(t)dB_t\right\| > K^*\right) \xrightarrow[T\to\infty]{} 0,$$

and

$$\mathrm{P}\left(\frac{1}{\sqrt{T}}\left|\int_0^T X(t)dB_t\right| > K^*\right) \xrightarrow[T\to\infty]{} 0,$$

which means

$$\frac{1}{\sqrt{T}}\|W_{(0,T)}(p)\| = O_p(\log^{a^*} T) \text{ and } \frac{1}{T}\|W_{(0,T)}(p)\|^2 = O_p(\log^{2a^*} T). \tag{A.112}$$

This completes the proof. □

A.4 Asymptotic distribution risk and relative efficiency

Let $\Lambda = \tilde{\Sigma}_{c^0}^{-1}(p_0) - \tilde{G}^*(c^0, p_0)\tilde{M}(c^0, p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)$, the following propositions, which establish the comparison of ADR between different estimators.

Proposition A.19. *If Assumption 2.1-2.3, Assumption 2.4 along with the set of local alternatives in* (2.4.16) *hold, then,* $\mathrm{ADR}\left(\hat{\theta}_T(\hat{s}, \hat{c}, \hat{p}), \boldsymbol{\theta}, \Omega\right) = \sigma^2 \mathrm{trace}(\Omega\tilde{\Sigma}_{c^0}^{-1}(p_0))$*; and*

$$\mathrm{ADR}\left(\tilde{\theta}_T(\hat{s}, \hat{c}, \hat{p}), \boldsymbol{\theta}, \Omega\right) = \mathrm{ADR}\left(\hat{\theta}_T(\hat{s}, \hat{p}), \boldsymbol{\theta}, \Omega\right)$$
$$-\mathrm{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0, p_0)\tilde{M}(c^0, p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right) + \tilde{r}_0^\top \tilde{G}^{*\top}(c^0, p_0)\Omega\tilde{G}^*(c^0, p_0)\tilde{r}_0.$$

The proof of this proposition follows from Proposition 2.3.20 along with some algebraic computations.

Proposition A.20. *If Assumption 2.1-2.3, Assumption 2.4 along with the set of local alternatives in* (2.4.16) *hold, then,*

$$\mathrm{ADR}\left(\hat{\theta}^{s}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)=\sigma^{2}\mathrm{trace}(\Omega\Lambda)+r_{0}^{\top}G^{*\top}(c^{0},p_{0})\Omega\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0}$$
$$-2\mathbb{E}\left[\gamma\left(\chi^{2}_{q+2}(\Delta)\right)\right]r_{0}^{\top}G^{*\top}(c^{0},p_{0})\Omega\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0}+\mathbb{E}\left[\gamma^{2}\left(\chi^{2}_{q+4}(\Delta)\right)\right]\tilde{r}_{0}^{\top}\tilde{G}^{*\top}(c^{0},p_{0})\times$$
$$\Omega\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0}+\mathbb{E}\left[\gamma^{2}\left(\chi^{2}_{q+2}(\Delta)\right)\right]\mathrm{trace}\left(\Omega\sigma^{2}\tilde{G}^{*}(c^{0},p_{0})\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\right).$$

Proof. By (3.4.3),

Proposition 2.3.20, Proposition A.9

$$\mathrm{ADR}\left(\hat{\theta}^{s}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)=\mathbb{E}\left[\mathrm{trace}\left(\tilde{\varrho}^{\top}\Omega\tilde{\varrho}\right)\right]+2\mathbb{E}\left[\gamma\left(\|\tilde{\varsigma}\|^{2}_{\Gamma}\right)\mathrm{trace}\left(\tilde{\varrho}^{\top}\Omega\tilde{\varsigma}\right)\right]$$
$$+\mathbb{E}\left[\gamma^{2}\left(\|\tilde{\varsigma}\|^{2}_{\Gamma}\right)\mathrm{trace}\left(\tilde{\varsigma}^{\top}\Omega\tilde{\varsigma}\right)\right].$$

By Theorem 3.1 in Nkurunziza [2012],

$$\mathrm{ADR}\left(\hat{\theta}^{s}(\hat{s},\hat{c},\hat{p}),\tilde{\theta},\Omega\right)=\mathrm{trace}\left(\Omega(\sigma^{2}\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})-\sigma^{2}\tilde{G}^{*}(c^{0},p_{0})\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0}))\right)$$
$$+\mathrm{trace}\left((\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})^{\top}\Omega(\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})\right)$$
$$-2\mathbb{E}\left[\gamma\left(\chi^{2}_{q+2}\left(\mathrm{trace}\left((\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})^{\top}\frac{1}{\sigma^{2}}\tilde{M}^{\top}(c^{0},p_{0})(\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\tilde{M}^{\top}(c^{0},p_{0}))^{-1}\times\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\tilde{M}(c^{0},p_{0})(\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})\right)\right)\right)\right]\mathrm{trace}\left((\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})^{\top}\Omega(\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})\right)$$
$$+\mathbb{E}\left[\gamma^{2}\left(\chi^{2}_{q+2}\left(\mathrm{trace}\left((\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})^{\top}\frac{1}{\sigma^{2}}\tilde{M}^{\top}(c^{0},p_{0})(\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\tilde{M}^{\top}(c^{0},p_{0}))^{-1}\times\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\tilde{M}(c^{0},p_{0})(\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})\right)\right)\right)\right]\mathrm{trace}\left(\Omega\sigma^{2}\tilde{G}^{*}(c^{0},p_{0})\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\right)$$
$$+\mathbb{E}\left[\gamma^{2}\left(\chi^{2}_{q+4}\left(\mathrm{trace}\left((\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})^{\top}\frac{1}{\sigma^{2}}\tilde{M}^{\top}(c^{0},p_{0})(\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\tilde{M}^{\top}(c^{0},p_{0}))^{-1}\times\right.\right.\right.\right.$$
$$\left.\left.\left.\left.\tilde{M}(c^{0},p_{0})(\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})\right)\right)\right)\right]\mathrm{trace}\left((\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})^{\top}\Omega(\tilde{G}^{*}(c^{0},p_{0})\tilde{r}_{0})\right).$$

It is obvious that $\mathrm{trace}\left(\Omega(\sigma^{2}\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})-\sigma^{2}\tilde{G}^{*}(c^{0},p_{0})\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0}))\right)=\sigma^{2}\mathrm{trace}(\Omega\Lambda)$. Since $\tilde{G}^{*}(c^{0},p_{0})=\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\tilde{M}^{\top}(c^{0},p_{0})(\tilde{M}(c^{0},p_{0})\tilde{\Sigma}^{-1}_{c^{0}}(p_{0})\tilde{M}^{\top}(c^{0},p_{0}))^{-1}$, and $\tilde{r}_{0}$ is a q column vector, $\operatorname{trace}\left((\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\Omega(\tilde{G}^*(c^0,p_0)\tilde{r}_0)\right) = \tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0$. Further,

$$\operatorname{trace}\left((\tilde{G}^*(c^0,p_0)\tilde{r}_0)^\top\frac{1}{\sigma^2}\tilde{M}^\top(c^0,p_0)(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0))^{-1}\tilde{M}(c^0,p_0)(\tilde{G}^*(c^0,p_0)\tilde{r}_0)\right)$$
$$= \frac{1}{\sigma^2}\left(r_0^\top(\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0))^{-1})^\top\tilde{M}^\top(\tilde{M}\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0))^{-1}\right.$$
$$\left.\times\,\tilde{M}(c^0,p_0)\,\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0))^{-1}\tilde{r}_0)\right)$$
$$= \frac{1}{\sigma^2}\left(\tilde{r}_0^\top(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0))^{-1}\tilde{r}_0\right) = \Delta.$$

Then,

$$\begin{aligned}\operatorname{ADR}\left(\hat{\theta}^s(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) &= \sigma^2\operatorname{trace}(\Omega\Lambda) + \tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\\ &-2\mathbb{E}\left[\gamma\left(\chi^2_{q+2}(\Delta)\right)\right]\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\\ &+\mathbb{E}\left[\gamma^2\left(\chi^2_{q+2}(\Delta)\right)\right]\operatorname{trace}\left(\Omega\sigma^2\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\\ &+\mathbb{E}\left[\gamma^2\left(\chi^2_{q+4}(\Delta)\right)\right]\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0.\end{aligned}$$

This completes the proof. □

If let the functions $\gamma(x) = \left(1-\frac{q-2}{x}\right)$, $x>0$, and $\gamma(x) = [1-(q-2)/x]\,\mathbb{I}_{\{x\geqslant q-2\}}$, $x>0$, the following propositions give the ADR of SEs.

Proposition A.21. *If Assumption 2.1-2.3, Assumption 2.4 along with the local alternative restriction* (2.4.16) *hold, then,*

$$\begin{aligned}&\operatorname{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) = \operatorname{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)\\ &+(q+2)(q-2)\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\chi^{-4}_{q+4}(\Delta)\right]\\ &-(q-2)\sigma^2\operatorname{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\left(2\mathbb{E}\left[\chi^{-2}_{q+2}(\Delta)\right]-(q-2)\mathbb{E}\left[\chi^{-4}_{q+2}(\Delta)\right]\right).\end{aligned}$$

Proof. Let $\gamma(x) = 1-\frac{q-2}{x}$, $x>0$. The proof follows from Proposition A.20. □

Proposition A.22. *If Assumption 2.1-2.3, Assumption 2.4 along with the local alternative restriction* (2.4.16) *hold, then,*

$$\operatorname{ADR}\left(\hat{\theta}_T^{sh+}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) = \operatorname{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)$$

$$+2\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)\mathbb{I}_{\{\chi^2_{q+2}(\Delta)<q-2\}}\right]$$

$$-\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi^2_{q+2}(\Delta)<q-2\}}\right]$$

$$-\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+4}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi^2_{q+4}(\Delta)<q-2\}}\right].$$

Proof. Let $\gamma(x) = (1-\frac{q-2}{x})^+, x>0$. The proof follows from Proposition A.20 along with some algebraic computations. □

Proof of Proposition 2.5.1. Let $f_0 = \text{ADR}\left(\tilde{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) - \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)$. By Proposition A.19,

$$f_0 = -\text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right)+\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0.$$

Note that the matrix $\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)$ is a $(c^0+1)(p_0+1)\times(c^0+1)(p_0+1)$ symmetric matrix, and

$$\tilde{G}^{*\top}(c^0,p_0)\Gamma\tilde{G}^*(c^0,p_0) = \left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\frac{1}{\sigma^2}\tilde{M}^\top(c^0,p_0)$$
$$\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}^{-1}\tilde{M}^\top(c^0,p_0)\right)^{-1}\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\times$$
$$\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)_{c^0}^{-1}(p_0)$$
$$=\frac{1}{\sigma^2}\left(\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\tilde{M}^\top(c^0,p_0)\right)^{-1},$$

which is positive definite for $\sigma>0$. By using Theorem 2.4.7 in Mathai and Provost [1992], and the fact that $\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Gamma\tilde{G}^*(c^0,p_0)\tilde{r}_0=\Delta$,

$$\lambda_{\min}\Delta \leqslant \tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0 \leqslant \lambda_{\max}\Delta, \tag{A.113}$$

which implies that

$$\lambda_{\min}\Delta-\text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right)\leqslant f_0$$
$$\leqslant \lambda_{\max}\Delta-\text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right).$$

Then, if $0\leqslant\lambda_{\min}\Delta-\text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right)$, i.e.

$\Delta \geqslant \text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right)/\lambda_{\min}$, $f_0 \geqslant 0$ this means

$$\text{ADR}\left(\tilde{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \geqslant \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right).$$

On the other hand, if $\lambda_{\max}\Delta - \text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right) \leqslant 0$, i.e. $\Delta \leqslant \text{trace}\left(\Omega\sigma^2\left(\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\right)/\lambda_{\max}$, one has $f_0 \leqslant 0$ which is equivalent to $\text{ADR}\left(\tilde{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)$. This completes the proof. □

Proof of Proposition 2.5.2. Let $f_1 = \text{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) - \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)$. By the identity [Saleh, 2006, page 32], $\Delta\mathbb{E}\left[\chi_{q+4}^{-4}\left(\tilde{\Delta}\right)\right] = \mathbb{E}\left[\chi_{q+2}^{-2}(\Delta)\right] - (q-2)\mathbb{E}\left[\chi_{q+2}^{-4}(\Delta)\right]$. Then, together with Proposition A.21

$$f_1 = \Big((q+2)(q-2)\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0$$

$$-2(q-2)\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\Delta\Big)\times$$

$$\mathbb{E}\left[\chi_{q+4}^{-4}(\Delta)\right] - (q-2)^2\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\mathbb{E}\left[\chi_{q+2}^{-4}(\Delta)\right].$$

Note that $\Delta = 0$ if and only if $\tilde{r}_0 = 0$. First, if $\Delta = 0$,

$$f_1 = -(q-2)^2\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\mathbb{E}\left[\chi_{q+2}^{-4}(\Delta)\right].$$

From the positive definite property of the matrix $\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)$,

$$\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right) \geqslant 0, \tag{A.114}$$

which implies that for $\Delta = 0$, $f_1 \leqslant 0$ which is equivalent to

$$\text{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) - \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant 0.$$

Further, for $\Delta > 0$, $\text{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) - \text{ADR}\left(\hat{\theta}_T(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant 0$, if $(q+2)(q-2)\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0 - 2(q-2)\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\Delta < 0$, which is

$$(q+2)(q-2)\tilde{r}_0^\top\tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0 \leqslant 2(q-2)\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\Delta.$$

Since $q > 2$, this is equivalent to

$$(q+2)\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0 \leqslant 2\sigma^2 \text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\tilde{\Delta}.$$

From (A.113), $f_1 \leqslant 0$ whenever $(q+2)\lambda_{\max}\Delta \leqslant 2\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\Delta$, which is $2\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)/\lambda_{\max} \geqslant (q+2)$. Further, let $f_2 = \text{ADR}\left(\hat{\theta}_T^{sh+}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) - \text{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right)$. From Proposition A.22,

$$\begin{aligned}
f_2 &= 2\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right]\\
&-\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right]\\
&-\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+4}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+4}^2(\Delta)<q-2\}}\right],
\end{aligned}$$

which implies that

$$f_2 = 2\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right]$$

$$-\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right]$$

$$-\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+4}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+4}^2(\Delta)<q-2\}}\right].$$

Note that $\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}} < 0$, thus,

$$\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right] < 0.$$

As a result of this fact,

$$2\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right] < 0.$$

Moreover, from

$$\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}} \geqslant 0 \text{ and } \left(1-(q-2)\chi_{q+4}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+4}^2(\Delta)<q-2\}} \geqslant 0,$$

together with (A.114) and the fact $\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0 \geqslant 0$,

$$-\sigma^2\text{trace}\left(\Omega\tilde{G}^*(c^0,p_0)\tilde{M}(c^0,p_0)\tilde{\Sigma}_{c^0}^{-1}(p_0)\right)\mathbb{E}\left[\left(1-(q-2)\chi_{q+2}^{-2}(\Delta)\right)^2\mathbb{I}_{\{\chi_{q+2}^2(\Delta)<q-2\}}\right] \leqslant 0$$

and $-\tilde{r}_0^\top \tilde{G}^{*\top}(c^0,p_0)\Omega\tilde{G}^*(c^0,p_0)\tilde{r}_0\mathbb{E}\left[(1-(q-2)\chi_{q+4}^{-2}(\Delta))\mathbb{I}_{\{\chi_{q+4}^2(\Delta)<q-2\}}\right] \leqslant 0$. This gives

that $\mathrm{ADR}\left(\hat{\theta}_T^{sh+}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) - \mathrm{ADR}\left(\hat{\theta}_T^{sh}(\hat{s},\hat{c},\hat{p}),\boldsymbol{\theta},\Omega\right) \leqslant 0$, for all $\Delta \geqslant 0$. This completes the proof. □

B Proofs related to GCIR process

B.1 On existence of the solution of the GCIR process

Proposition B.1. *Proposition 2.13 (Yamada & Watanabe (1971)) Karatzas and Shreve [1998]. Let us suppose that the coefficients of the one-dimensional equation (d=r=1),*

$$dX(t) = \varphi(t, X(t))dt + \sigma(t, X(t))dW_t \tag{B.1}$$

satisfy the condition: $|\varphi(t,x)-\varphi(t,y)| \leqslant K|x-y|, \quad |\sigma(t,x)-\sigma(t,y)| \leqslant h(|x-y|)$, *for every* $0 \leqslant t < \infty$ *and* $x \in \mathbb{R}, y \in \mathbb{R}$, *where K is a positive constant and* $h : [0,\infty) \to [0,\infty)$ *is a strictly increasing function with* $h(0) = 0$ *and* $\int_{(0,\epsilon)} h^{-2}(u)du = \infty;$ for any $\epsilon > 0$. *Then strong and uniqueness hold for the solution of the stochastic differential equation (B.1).*

Proposition B.2. *Proposition 2.18 (Karatzas and Shreve [1998], pp.293) Suppose that on a certain probability space* $(\Omega, \mathcal{F}, \mathbb{P})$ *equipped with a filtration* $\{\mathcal{F}_t\}$ *which satisfies the usual conditions, a standard, one-dimensional Brownian motion* $\{B_t, \mathcal{F}_t, 0 \leqslant t < \infty\}$ *and two adapted processes* $X^{(j)}(t)$, $j = 1,2$, *such that* $dX^{(j)}(t) = X_0^{(j)} + \int_0^t b_j(s, X^{(j)}(s))ds + \int_0^t \sigma(s, X^{(j)}(s))dB_s;\ 0 \leqslant t < \infty$ *holds a.s. for* $j = 1,2$. *Suppose that the following statements hold.*

1. *The coefficients* $\sigma(t,x), b_j(t,x)$ *are continuous, real-valued functions on* $[0,+\infty)\times\mathbb{R}$,
2. *The dispersion matrix* $\sigma(t,x)$ *satisfies* $|\sigma(t,x)-\sigma(t,y)| \leqslant h(|x-y|)$ *where* $h : [0,+\infty) \to [0,+\infty)$ *is strictly increasing with* $h(0) = 0$ *and* $\int_0^\varepsilon h^{-2}(u)du = \infty;\ \forall\, \varepsilon > 0$,

3. $X^{(1)}(0) \leqslant X^{(2)}(0)$ *a.s.*

4. $b_1(t,x) \leqslant b_2(t,x),\ \forall\ 0 \leqslant t < \infty,\ x \in \mathbb{R}$,

5. *Both* $b_1(t,x)$ *and* $b_2(t,x)$ *satisfy the Lipschitz condition in x.*

Then, $\mathrm{P}\left(X^{(1)}(t) \leqslant X^{(2)}(t),\ \forall\ 0 \leqslant t < \infty\right) = 1.$

Proof of Proposition 3.2.1. By the Proposition 2.13 (*Yamada-Watanabe Theorem* 1971) Karatzas and Shreve [1998], all requirements are checked: ($\forall\ x, y \geqslant 0$).

(1) First, by (3.2.1),

$|S(t,x) - S(t,y)| = |(L(t) - \alpha x) - (L(t) - \alpha y)| = \alpha|x-y|$, which implies that the drift term $S(t,x)$ satisfies Lipschitz condition.

(2) Second, the function $\sigma\sqrt{x}$ vanishes at 0 and satisfies the Hölder condition, which is $\left|\sigma\sqrt{x} - \sigma\sqrt{y}\right| \leqslant \sigma\sqrt{|x-y|}$. Indeed, choose $H(u) = \sigma u^{1/2}$ and note that

$$\sqrt{|x-y|} - |\sqrt{x} - \sqrt{y}| = \begin{cases} \sqrt{|(\sqrt{x} - \sqrt{y})| \frac{2\sqrt{y}}{\sqrt{(\sqrt{x}+\sqrt{y})} + \sqrt{|\sqrt{x}-\sqrt{y}|}}} \geqslant 0, & \text{if } x \geqslant y \geqslant 0; \\ \sqrt{|(\sqrt{x} - \sqrt{y})| \frac{2\sqrt{x}}{\sqrt{(\sqrt{x}+\sqrt{y})} + \sqrt{|\sqrt{x}-\sqrt{y}|}}} \geqslant 0, & \text{if } 0 \leqslant x \leqslant y. \end{cases}$$

Then,

$$|\sqrt{x} - \sqrt{y}|^2 \leqslant |x-y|. \tag{B.2}$$

This implies that $|\sigma(t,x) - \sigma(t,y)| = \sigma|\sqrt{x} - \sqrt{y}| \leqslant \sigma\sqrt{|x-y|} = h(|x-y|)$. It is clear that $\sqrt{|x-y|} - |\sqrt{x} - \sqrt{y}| = 0$, if $x = 0$ or $y = 0$ or $x = y$, otherwise, $\sqrt{|x-y|} - |\sqrt{x} - \sqrt{y}| > 0$.

By the Proposition 2.13 (*Yamada-Watanabe Theorem* 1971) Karatzas and Shreve [1998], the SDE (3.2.1) admits a strong and unique solution in $[0,\infty)$. Further, by Proposition B.2 and Assumption 3.1, that $0 \leqslant X(t) \leqslant X(t)$ a.s. This completes the proof. □

Proof of Corollary 3.2.1. By (3.2.1), $S(\theta, t, X(t)) = L(t) - \alpha X(t)$, and $L(t) = \sum_{i=1}^{p} \mu_i \varphi_i(t)$. Then, together with Assumption 2.1-2.2, Assumption 3.1-3.2,

$$\left(\frac{S(\theta, t, X(t))}{\sigma \sqrt{X(t)}}\right)^2 < L^2(t)/(\sigma^2 X(t)) + X(t)/\sigma^2,$$

where, from Assumption 3.2 and (B.32), $\sup_{t \geqslant 0} \mathbb{E}[1/X(t)] < \infty$. Further, (3.2.4) implies that $\sup_{t \geqslant 0} \mathbb{E}[X(t)] < \infty$, which gives $\mathbb{E}\left[\int_0^T \left(S^2(\theta, t, X(t))/(\sigma^2 X(t))\right) dt\right] < \infty$. Then $\mathbb{P}\left(\int_0^T \left(S(\theta, t, X(t))/(\sigma \sqrt{X(t)})\right)^2 dt < \infty\right) = 1$, for all $0 \leqslant T < \infty$. This completes the proof. □

Proof of Proposition 3.2.2. **(1)**.Let $G(t, X(t)) = e^{\alpha t} X(t)$, by Itô's Lemma,

$$dG(t, X(t)) = \frac{\partial G}{\partial t}(t, X(t))dt + \frac{\partial G}{\partial X(t)}(t, X(t))dX(t) + \frac{1}{2}\frac{\partial^2 G}{\partial X^2(t)}(t, X(t))d\langle X(t), X(t)\rangle,$$

where $\langle \cdot, \cdot \rangle$ denote the variation. Since $dX(t) = (L(t) - \alpha X(t))dt + \sigma \sqrt{X(t)} dB_t$, then

$$dG(t, X(t)) = \alpha e^{\alpha t} X(t)dt + e^{\alpha t} dX(t) = \alpha e^{\alpha t} X(t)dt + e^{\alpha t}\left((L(t) - \alpha X(t))dt + \sigma \sqrt{X(t)} dB_t\right),$$

which implies that $d(e^{\alpha t} X(t)) = e^{\alpha t} L(t)dt + \sigma e^{\alpha t} \sqrt{X(t)} dB_t$. Integrating both sides from 0 to t implies that

$$e^{\alpha t} X(t) = X_0 + \int_0^t e^{\alpha s} L(s)ds + \int_0^t \sigma e^{\alpha s} \sqrt{X(s)} dB_s.$$

Then

$$X(t) = e^{-\alpha t} X_0 + e^{-\alpha t} \int_0^t e^{\alpha s} L(s)ds + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} \sqrt{X(s)} dB_s.$$

This completes the proof of Part **(1)**.

(2). Since Itô integrals are local martingales (see Theorem 13.2 in Klebaner [2005]), $\int_0^t \sqrt{X(s)} dB_s$ is a local martingale. For $n \in \mathbb{N}_+$, define a stopping time $T_n(\omega)$ by $T_n(\omega) := \inf\{t \geqslant 0, X(t, \omega) \geqslant n\}$. let $T_n = T_n(\omega)$. Since $\{X(t), t \geqslant 0\}$ has almost surely continuous sample paths, it holds that $X(t \wedge T_n) \leqslant n$. So that $\int_0^{t \wedge T_n} \sqrt{X(s)} dB_s$ is a martingale in t for any fixed n, then from (3.2.1),

$$X(t \wedge T_n) = X_0 + \int_0^{t\wedge T_n} (L(s) - \alpha X(s))\, ds + \sigma \int_0^{t\wedge T_n} \sqrt{X(s)} dB_s. \tag{B.3}$$

Taking expectation both sides,

$$\mathbb{E}[X(t \wedge T_n)] = \mathbb{E}[X_0] + \mathbb{E}\left[\int_0^{t\wedge T_n} (L(s) - \alpha X(s) ds\right]. \tag{B.4}$$

Since $X(t)$ is non-negative, $\int_0^{t\wedge T_n(\omega)} X(s,\omega)ds$ is increasing and $\lim_{n\to\infty} \int_0^{t\wedge T_n(\omega)} X(s,\omega)ds = \int_0^t X(s,\omega)ds$ for all $\omega \in \Omega$, $\int_0^{t\wedge T_n} X(s)ds \xrightarrow[n\to\infty]{a.s.} \int_0^t X(s)ds$. Therefore, by monotone convergence theorem, $\int_0^{t\wedge T_n} X(s)ds \xrightarrow[n\to\infty]{L^1} \int_0^t X(s)ds$, which implies that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^{t\wedge T_n} X(s)ds\right] = \mathbb{E}\left[\int_0^t X(s)ds\right]. \tag{B.5}$$

Similarly, $\int_0^{t\wedge T_n} L(s)ds \xrightarrow[n\to\infty]{a.s.} \int_0^t L(s)ds$, so, $\lim_{n\to\infty} \mathbb{E}\left[X(t\wedge T_n)\right] = \mathbb{E}[X_0] + \mathbb{E}[\int_0^t (L(s) - \alpha X(s))ds]$. Since $X(t\wedge T_n) \xrightarrow[n\to\infty]{a.s.} X(t)$, by using Fatou's Lemma,

$$\begin{aligned}
\mathbb{E}[X(t)] = \mathbb{E}[\liminf_{n\to\infty} X(t\wedge T_n)] &\leqslant \liminf_{n\to\infty} \mathbb{E}[X(t\wedge T_n)] = \lim_{n\to\infty} \mathbb{E}[X(t\wedge T_n)] \\
&= \mathbb{E}[X_0] + \int_0^t L(s)ds - \alpha \mathbb{E}\left[\int_0^t X(s)ds\right].
\end{aligned}$$

Since $\alpha > 0$, $\mathbb{E}\left[\int_0^t X(s)ds\right] > 0$, $\mathbb{E}[X(t)] < \mathbb{E}[X_0] + \int_0^t L(s)ds$. By using (3.2.3),

$$\mathbb{E}[X(t)] < \mathbb{E}[X_0] + K_\mu K_\varphi t. \tag{B.6}$$

It follows that the quadratic variation is

$$\left\langle \int_0^{\cdot} e^{-\alpha(t-s)} \sigma \sqrt{X(s)} dB_s \right\rangle_t = \int_0^t e^{-2\alpha(t-s)} \sigma^2 X(s)ds,$$

and by (B.6),

$$\begin{aligned}
\mathbb{E}\left[\int_0^t e^{-2\alpha(t-s)}\sigma^2 X(s)ds\right] &= \int_0^t e^{-2\alpha(t-s)}\sigma^2 \mathbb{E}\left[X(s)\right] ds \\
&< \int_0^t e^{-2\alpha(t-s)}\sigma^2 \left(\mathbb{E}[X_0] + K_\mu K_\varphi s\right) ds.
\end{aligned}$$

Further,

$$\int_0^t e^{-2\alpha(t-s)}\sigma^2\left(\mathbb{E}[X_0]\right)ds = \sigma^2\mathbb{E}[X_0]\int_0^t e^{-2\alpha(t-s)}ds = \sigma^2\mathbb{E}[X_0]\frac{1}{2\alpha}(1-e^{-2\alpha t}),$$

and

$$\int_0^t e^{-2\alpha(t-s)}\sigma^2\left(K_\mu K_\varphi s\right)ds = \sigma^2 K_\mu K_\varphi\int_0^t e^{-2\alpha(t-s)}sds = \sigma^2 K_\mu K_\varphi\frac{1}{2\alpha}\left(t-\frac{1}{2\alpha}\left(1-e^{-2\alpha t}\right)\right).$$

This implies that

$$\mathbb{E}\left[\int_0^t e^{-2\alpha(t-s)}\sigma^2 X(s)ds\right] < \sigma^2\mathbb{E}[X_0]\frac{1}{2\alpha}\left(1-e^{-2\alpha t}\right)+\sigma^2 K_\mu K_\varphi\frac{1}{2\alpha}\left(t-\frac{1}{2\alpha}\left(1-e^{-2\alpha t}\right)\right),$$

which gives

$$\mathbb{E}\left[\left\langle\int_0^{\cdot} e^{-\alpha(t-s)}\sigma\sqrt{X(s)}dB_s\right\rangle_t\right] < +\infty \tag{B.7}$$

Thus, the term $\int_0^t e^{-\alpha(t-s)}\sigma\sqrt{X(s)}dB_s$ is a martingale. Then, from (3.2.2),

$$\mathbb{E}[X(t)] = e^{-\alpha t}\mathbb{E}[X_0]+e^{-\alpha t}\int_0^t e^{\alpha s}L(s)ds = e^{-\alpha t}\mathbb{E}[X_0]+\sum_{i=1}^p\mu_i\int_0^t e^{-\alpha(t-s)}\varphi_i(s)ds. \tag{B.8}$$

This completes the proof of the second assertion.

(3) From Assumption 3.1 and (3.2.2),

$$\mathbb{E}[X^2(t)] = e^{-2\alpha t}\mathbb{E}\left[X_0^2\right]+e^{-2\alpha t}\left(\int_0^t e^{\alpha s}L(s)ds\right)^2+\sigma^2 e^{-2\alpha t}\mathbb{E}\left[\left(\int_0^t e^{\alpha s}\sqrt{X(s)}dB_s\right)^2\right]$$
$$+2e^{-2\alpha t}\mathbb{E}\left[X_0\right]\int_0^t e^{\alpha s}L(s)ds+2\sigma e^{-2\alpha t}\int_0^t e^{\alpha s}L(s)ds\,\mathbb{E}\left[\int_0^t e^{\alpha s}\sqrt{X(s)}dB_s\right]$$
$$+2e^{-\alpha t}\mathbb{E}[X_0]\,\sigma e^{-\alpha t}\mathbb{E}\left[\int_0^t e^{\alpha s}\sqrt{X(s)}dB_s\right].$$

By Itô isometry and (B.8),

$$\mathbb{E}\left[\left(\int_0^t e^{\alpha s}\sqrt{X(s)}dB_s\right)^2\right] = \mathbb{E}[X_0]\frac{1}{\alpha}(e^{\alpha t}-1)+\int_0^t e^{\alpha s}\int_0^s e^{\alpha u}L(u)duds.$$

From (B.7), $\int_0^t e^{\alpha s}\sqrt{X(s)}dB_s$ is a martingale, then, $\mathbb{E}\left[\int_0^t e^{\alpha s}\sqrt{X(s)}dB_s\right] = 0$. Hence,

$$\mathbb{E}[X^2(t)] = e^{-2\alpha t}\mathbb{E}\left[X_0^2\right]+e^{-2\alpha t}\left(\int_0^t e^{\alpha s}L(s)ds\right)^2+\sigma^2 e^{-2\alpha t}\Bigg(\mathbb{E}[X_0]\frac{1}{\alpha}(e^{\alpha t}-1)$$

$$+\int_0^t e^{\alpha s}\int_0^s e^{\alpha u}L(u)duds\Big) + 2e^{-2\alpha t}\mathbb{E}[X_0] \int_0^t e^{\alpha s}L(s)ds,$$

which implies that

$$\mathrm{Var}(X(t)) = e^{-2\alpha t}\left(\mathrm{Var}(X_0) + \sigma^2\mathbb{E}[X_0]\frac{1}{\alpha}(e^{\alpha t}-1) + \sigma^2\int_0^t e^{\alpha s}\int_0^s e^{\alpha u}L(u)duds\right).$$

Further, by combining (3.2.3) and the fact that $\sup\limits_{t\geqslant 0}(e^{-\alpha t} - e^{-2\alpha t}) = \frac{1}{4}$,

$$\sup_{t\geqslant 0}\mathbb{E}[X^2(t)] \leqslant \mathbb{E}\left[X_0^2\right] + \frac{K_\varphi^2 K_\mu^2}{\alpha^2} + \sigma^2\mathbb{E}[X_0]\frac{1}{4\alpha} + \sigma^2 K_\varphi K_\mu\left(\frac{1}{2\alpha^2} - \frac{1}{4\alpha^2}\right) + K_\varphi K_\mu\frac{1}{2\alpha},$$

which implies that $\sup\limits_{t\geqslant 0}\mathbb{E}[X^2(t)] < \infty$. This completes the proof. □

Proof of Proposition 3.2.3. The case where $m = 2$ is proven in Proposition 3.2.2. Thus, in this proposition, only $m > 2$ is considered. Let $Y(t, X(t)) = e^{\alpha m t}X^m(t)$, by Itô's formula,

$$\begin{aligned} dY(t,X(t)) =& \alpha m e^{\alpha m t}X^m(t)dt + m e^{\alpha m t}X^{m-1}(t)dX(t) + \frac{1}{2}m(m-1)e^{\alpha m t}(X(t))^{m-2}d\langle X(t),X(t)\rangle \\ =& \alpha m e^{\alpha m t}X^m(t)dt + m e^{\alpha m t}X^{m-1}(t)\left(L(t) - \alpha X(t)\right)dt + m e^{\alpha m t}X^{m-1}(t)\sigma\sqrt{X(t)}dB_t \\ &+ \frac{1}{2}m(m-1)e^{\alpha m t}(X(t))^{m-2}d\langle X(t),X(t)\rangle \\ =& m e^{\alpha m t}X^{m-1}(t)\left(L(t) + \sigma^2\frac{1}{2}(m-1)\right)dt + \sigma m e^{\alpha m t}(X(t))^{m-1/2}dB_t. \end{aligned}$$

Then,

$$e^{\alpha m t}X^m(t) = X_0^m + \int_0^t m e^{\alpha m s}X^{m-1}(t)\left(L(s) + \sigma^2\frac{1}{2}(m-1)\right)ds + \int_0^t \sigma m e^{\alpha m s}(X(t))^{m-1/2}dB_s,$$

which implies that

$$\begin{aligned} X^m(t) = e^{-\alpha m t}X_0^m + m\int_0^t e^{-\alpha m(t-s)}X^{m-1}(t)\left(L(s) + \sigma^2\frac{1}{2}(m-1)\right)ds \\ + \sigma m\int_0^t e^{-\alpha m(t-s)}(X(t))^{m-1/2}dB_s, \end{aligned}$$

Let $\tau_n = \inf\{t > 0 : X(t) \geqslant n\}$, then,

$$X^m(t \wedge \tau_n) = e^{-\alpha m t \wedge \tau_n} X_0^m + m \int_0^{t \wedge \tau_n} e^{-\alpha m(t \wedge \tau_n - s)} X^{m-1}(t) \left(L(s) + \sigma^2 \frac{1}{2}(m-1) \right) ds$$
$$+\sigma m \int_0^{t \wedge \tau_n} e^{-\alpha m(t \wedge \tau_n - s)} (X(t))^{m-1/2} dB_s. \tag{B.9}$$

Since $M_t = \int_0^t e^{\alpha m s}(X(t))^{m-1/2} dB_s$ is an Itô integral, M_t is a local martingale. The stopping time τ_n is a localizing sequence, so that $M_{t\wedge\tau_n} := \int_0^{t\wedge\tau_n} e^{\alpha m s}(X(t))^{m-1/2} dB_s$ is a martingale and $\mathbb{E}[M_{t\wedge\tau_n}] = \mathbb{E}[M_0] = 0$. Since $X(t)$ is non-negative, $\int_0^{t\wedge T_n(\omega)} X^{m-1}(t)(\omega) ds$ is increasing and $\lim_{n\to\infty} \int_0^{t\wedge T_n(\omega)} X^{m-1}(t)(\omega) ds = \int_0^t X^{m-1}(t)(\omega) ds$ for all $\omega \in \Omega$. This implies that $\int_0^{t\wedge T_n} X^{m-1}(t)(\omega) ds \xrightarrow[n\to\infty]{a.s.} \int_0^t X^{m-1}(t)(\omega) ds$. Therefore, by monotone convergence theorem, $\int_0^{t\wedge T_n} X^{m-1}(t)(\omega) ds \xrightarrow[n\to\infty]{L^1} \int_0^t X^{m-1}(t)(\omega) ds$, which gives

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^{t\wedge T_n} X^{m-1}(t) ds \right] = \mathbb{E}\left[\int_0^t X^{m-1}(t) ds \right].$$

Further, since $X^m(t \wedge T_n) \xrightarrow[n\to\infty]{a.s.} X^m(t)$, taking expectation on both sides of (B.9) and by Fatou's Lemma,

$$\mathbb{E}[X^m(t)] = \mathbb{E}[\lim_{n\to\infty} X^m(t \wedge \tau_n)] = \mathbb{E}[\liminf_{n\to\infty} X^m(t \wedge \tau_n)]$$

$$\leqslant \liminf_{n\to\infty} \mathbb{E}[X^m(t \wedge \tau_n)] = \lim_{n\to\infty} \mathbb{E}[X^m(t \wedge \tau_n)].$$

Then,

$$\mathbb{E}[X^m(t)] \leqslant e^{-\alpha m t} \mathbb{E}[X_0^m] + m \left(K_\mu K_{t^*} + \frac{1}{2}\sigma^2(m-1) \right) e^{-\alpha m t} \int_0^t e^{\alpha m s} \mathbb{E}[X^{m-1}(t)] ds,$$

which gives

$$\mathbb{E}[X^m(t)] \leqslant \mathbb{E}[X_0^m] + C_m e^{-\alpha m t} \int_0^t e^{\alpha m s} \mathbb{E}[X^{m-1}(t)] ds, \tag{B.10}$$

where $C_m = m\left(K_\mu K_\varphi + \frac{1}{2}\sigma^2(m-1)\right)$. Then, by recursion of (B.10)

$$\mathbb{E}[X^m(t)] \leqslant \mathbb{E}[X_0^m] + C_m e^{-\alpha m t} \int_0^t e^{\alpha m s_1} \Big(\mathbb{E}[X_0^{m-1}] +$$

$$C_{m-1} e^{-\alpha(m-1)s_1} \int_0^{s_1} e^{\alpha(m-1)s_2} \mathbb{E}[X^{m-2}(s_2)] ds_2 \Big) ds_1.$$

Then,

$$\mathbb{E}[X^m(t)] \leqslant \mathbb{E}[X_0^m]+C_m\mathbb{E}[X_0^{m-1}]\frac{1}{\alpha m}+C_mC_{m-1}e^{-\alpha mt}\int_0^t e^{\alpha s_1}\left(\int_0^{s_1} e^{\alpha(m-1)s_2}\mathbb{E}[X^{m-2}(s_2)]ds_2\right)ds_1.$$

This gives

$$\begin{aligned}\mathbb{E}[X^m(t)] \leqslant &\mathbb{E}[X_0^m] + C_m\mathbb{E}[X_0^{m-1}]\frac{1}{\alpha m} + C_mC_{m-1}\mathbb{E}[X_0^{m-2}]\frac{1}{\alpha(m-1)}\frac{1}{\alpha m} + \cdots\\ &+ C_mC_{m-1}\cdots C_{m-(k-1)}\mathbb{E}\left[X_0^{m-(k-1)}\right]e^{-\alpha mt}\times\\ &\int_0^t e^{\alpha s_1}\int_0^{s_1} e^{\alpha s_2}\cdots\int_0^{s_{k-1}} e^{\alpha(m-(k-1))s_k}\mathbb{E}[X^{m-k}(s_k)]ds_k\cdots ds_2ds_1,\end{aligned}$$

for $0 < m-k \leqslant 1$. By Jensen's inequality,

$$\mathbb{E}[(X(t))^{m-k}] \leqslant \{\mathbb{E}[X(t)]\}^{m-k} \leqslant \begin{cases}1, & if\ \mathbb{E}[X(t)] \leqslant 1\\ \mathbb{E}[X(t)], & else.\end{cases}$$

This implies that

$$\begin{aligned}\mathbb{E}[X^m(t)] \leqslant \mathbb{E}[X_0^m] + C_m\mathbb{E}[X_0^{m-1}]\frac{1}{\alpha m} + C_mC_{m-1}\mathbb{E}[X_0^{m-2}]\frac{1}{\alpha(m-1)}\frac{1}{\alpha m} + \cdots\\ + \max\{1,\mathbb{E}[X(t)]\}C_mC_{m-1}\cdots C_{m-(k-1)}\mathbb{E}\left[X_0^{m-(k-1)}\right]\frac{1}{\alpha(m-(k-1))}\cdots\frac{1}{\alpha m}.\end{aligned}$$

Then, (3.2.6) holds where K_m is given by (3.2.5). This completes the proof. □

B.2 Auxiliary processes and approximate stationary processes

Theorem B.1 (Dubins-Schwarz theorem)**.** *Revuz and Yor [1999] Let $\boldsymbol{M}$ be a continuous local martingale with respect to a filtration $(\mathscr{F}_t)_{t\geqslant 0}$, such that $\boldsymbol{M}_0 = 0$ and $\langle \boldsymbol{M}\rangle_\infty = \infty$ almost surely. For all $t \geqslant 0$, let $\boldsymbol{T}_t = \inf\{s \geqslant 0 : \langle \boldsymbol{M}\rangle_s > t\} = \langle \boldsymbol{M}\rangle_t^{-1}$ be the generalized inverse of the non-decreasing process $\langle \boldsymbol{M}\rangle$ issued from 0. Then*

1. $\boldsymbol{B} = (\boldsymbol{M}_{\langle \boldsymbol{M}\rangle_t^{-1}})$ *is a Brownian motion with respect to the filtration* $(\mathscr{F}_{T_t})_{t\geqslant 0}$,
2. $(\boldsymbol{B}_{\langle \boldsymbol{M}\rangle_t})_{t\geqslant 0} = (\boldsymbol{M}_t)_{t\geqslant 0}$.

The following theorem is an extension of Dubins and Schwarz's (1965) theorem. It allows time change in terms of strictly increasing homeomorphisms $\phi(t)$ rather than just t. The proof has been given in Maghsoodi [1993].

Theorem B.2 (Extension of Dubins and Schwarz's (1965) theorem)**.** *Let $\boldsymbol{M}$ be a continuous local martingale with respect to a filtration $(\mathscr{F}_t)_{t\geqslant 0}$, such that $\boldsymbol{M}_0 = 0$ and $\langle \boldsymbol{M}\rangle_\infty = \infty$ almost surely. For all $t \geqslant 0$, let $\phi(t) = \int_0^t g^2(s)ds$ be a homeomorphism on $0 \leqslant t < \infty$ such that $\lim_{t\to\infty}\phi(t) = \infty$, define the stopping time $\tau(t) = \inf\{s \geqslant 0 : \langle \boldsymbol{M}\rangle_s > \phi(t)\}$. Then, the time changed process $N(t) := \boldsymbol{M}(\tau(t)), 0 \leqslant t < \infty$ is a $(\Omega, \mathscr{G}_t, \mathrm{P})$ square-integrable martingale with $\mathscr{G}_t = \mathscr{F}_{\tau(t)}$ and can be represented as $N(t) = \int_0^t g(s)d\boldsymbol{B}_s$ almost surely, where $\boldsymbol{B}_t$ is a $\mathscr{G}_t$ measurable Brownian motion.*

Theorem B.3 (Lévy's Charaterization of Brownian motion, Theorem 4.6.4 in Shreve [2004])**.** *Let $M(t)$, $t \geqslant 0$ be a martingale relative to a filtration $\mathscr{F}_t, t \geqslant 0$. Assume that $M(0) = 0$, $M(t)$ has continuous paths and $\langle M\rangle_t = t$ for all $t \geqslant 0$. Then, $M(t)$ is a standard Brownian motion.*

B.2.1 The case where the dimension is positive integer

Under this condition, the processes $\{X(t), t \geqslant 0\}$ is a sum of independent squared O-U process. Let $W_1, W_2, \cdots, W_d$ be independent Brownian motions. For $j = 1, 2, \cdots, d$, let $Y_j(t)$ be the O–U process

$$dY_j(t) = -\frac{\alpha}{2}Y_j(t)dt + \frac{1}{2}\sigma dW_j(t).$$

By Itô's formula,

$$Y_j(t) = e^{-\frac{\alpha}{2}t}Y_j(0) + \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}dW_j(s).$$

Set $X(t) = \sum\limits_{j=1}^{d} Y_j^2(t)$, with $X_0 = \sum\limits_{j=1}^{d} Y_j^2(0)$. Below, it is proven that $X(t)$ is the solution to the GCIR model (3.2.1). Indeed, by Itô's formula,

$$dX(t) = \sum_{j=1}^{d}\Big(2Y_j(t)dY_j(t) + d\langle Y_j(t)\rangle\Big) = \sum_{j=1}^{d}\Bigg(2Y_j(t)(-\frac{\alpha}{2}Y_j(t)dt + \frac{1}{2}\sigma dW_j(t)) + \frac{\sigma^2}{4}dt\Bigg).$$

Then,

$$dX(t) = \sum_{j=1}^{d}\left(\left(-\alpha Y_j(t) + \frac{\sigma^2}{4}\right)dt + \sigma Y_j(t)dW_j(t)\right) = (L(t) - \alpha X(t))\,dt + \sum_{j=1}^{d}\sigma Y_j(t)dW_j(t).$$

Since that $W_j, j = 1, 2, \cdots, d$ are independent Brownian motions, $\langle W_i, W_j\rangle_t = t\,\mathbb{I}_{\{i=j\}}$. Hence, the quadratic variation of B is given by

$$\langle B\rangle_t = \left\langle \sum_{j=1}^{d}\int_0^{\cdot}\frac{Y_j(s)}{\sqrt{X(s)}}dW_j(s), \sum_{j=1}^{d}\int_0^{\cdot}\frac{Y_j(s)}{\sqrt{X(s)}}dW_j(s)\right\rangle_t = t.$$

It follows from the Lévy characterization for Brownian motion [Shreve, 2004, Theorem B.3] that B is a standard Brownian motion. Then,

$$dX(t) = (L(t) - \alpha X(t))\,dt + \sigma\sqrt{X(t)}dB_t,$$

which shows that $X(t) = \sum_{j=1}^{d} Y_j^2(t)$ is a GCIR process.

From the definition of the process $\tilde{Y}_j(t)$ in Section 3.3.2, for $s \in (-\infty, 0)$, $\tilde{B}_j(s) = \bar{B}_j(-s)$,

$$\tilde{Y}_j(t) = \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}d\bar{B}_j(-s) + \int_0^t e^{\frac{\alpha}{2}s}dB_j(s)\right).$$

First, to establish the distribution of $\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}d\bar{B}_j(-s)$, let $u = -s$, $\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}d\bar{B}_j(-s) = -\int_0^{\infty} e^{-\frac{\alpha}{2}u}d\bar{B}_j(u)$. let $\mathcal{I}_U = \int_0^U e^{-\frac{\alpha}{2}u}d\bar{B}_j(u)$. Since the integrator $e^{-\frac{\alpha}{2}u}$ is not random, the Itô integral $\mathcal{I}_U$ follows a normal distribution with mean 0 and variance $\int_0^U e^{-\alpha u}du$. One can verify that

$$\int_{-\infty}^{0} e^{\alpha s}d\bar{B}_{-s} \sim \mathcal{N}\left(0, \frac{1}{\alpha}\right), \tag{B.11}$$

and then, by the independence between $\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}d\bar{B}_j(-s)$ and $\int_0^t e^{\frac{\alpha}{2}s}dB_j(s)$,

$$\tilde{Y}_j(t) = \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}d\bar{B}_j(-s) + \int_0^t e^{\frac{\alpha}{2}s}dB_j(s)\right) \sim \mathcal{N}\left(0, \frac{\sigma^2}{4}\frac{1}{\alpha}\right).$$

Let $\tilde{X}(t) = \sum_{j=1}^{d}\tilde{Y}_j^2(t)$ and note that $\tilde{Y}_1(t), \tilde{Y}_2(t), \ldots, \tilde{Y}_d(t)$ are independent. This implies that $\tilde{X}(t) = \frac{\sigma^2}{4\alpha}\sum_{j=1}^{d}\frac{4\alpha}{\sigma^2}\tilde{Y}_j^2(t) \sim \frac{\sigma^2}{4\alpha}\chi_d^2$ and the process $\{\tilde{X}(t), t \geqslant 0\}$ is strictly stationary and ergodic. Further, The following proposition shows that the distance between $X(t)$ and

$\tilde{X}(t)$ converges to 0 in L^m and almost surely. Further, the distance between $\frac{1}{X(t)}$ and $\frac{1}{\tilde{X}(t)}$ converges to 0 almost surely too.

Proposition B.3. *If Assumption 2.1-2.2, 3.1-3.2 hold, then, (1).* $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{a.s.} 0$*; (2).* $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{L^m} 0$*; (3).* $\dfrac{1}{\tilde{X}(t)} - \dfrac{1}{X(t)} \xrightarrow[t\to\infty]{a.s.} 0$.

Proof. **(1).** From the definition of $Y_j(t)$ and $\tilde{Y}_j(t)$,
$|Y_j(t) - \tilde{Y}_j(t)|^2 \leqslant 2\left(Y_j^2(0) + \frac{1}{4}\sigma^2\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\tilde{B}_j(s)\right)^2\right)e^{-\alpha t} = A_j e^{-\alpha t}$ with

$$\leqslant 2\left(e^{-\alpha t}Y_j^2(0) + \frac{1}{4}\sigma^2 e^{-\alpha t}\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\tilde{B}_j(s)\right)^2\right) = 2\left(Y_j^2(0) + \frac{1}{4}\sigma^2\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\tilde{B}_j(s)\right)^2\right)e^{-\alpha t}.$$

Let $A_j = 2\left(Y_j^2(0) + \frac{1}{4}\sigma^2\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\tilde{B}_j(s)\right)^2\right)$. Then,

$$\mathbb{E}[A_j] = 2\left(\mathbb{E}[Y_j^2(0)] + \frac{1}{4}\sigma^2\mathbb{E}\left[\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\tilde{B}_j(s)\right)^2\right]\right).$$

Since for $s \in (-\infty, 0)$, $\tilde{B}_s = \bar{B}_{-s}$,

$$\mathbb{E}\left[\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\tilde{B}_j(s)\right)^2\right] = \mathbb{E}\left[\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s}d\bar{B}_j(s)\right)^2\right] = \mathbb{E}\left[\left(\int_0^\infty e^{-\frac{\alpha}{2}u}d\bar{B}_j(u)\right)^2\right]. \tag{B.12}$$

Since $\mathcal{I}_U = \int_0^U e^{-\frac{\alpha}{2}u}d\bar{B}_j(u)$. By Itô's isometry,

$$\mathbb{E}[\mathcal{I}_U^2] = \mathbb{E}\left[\left(\int_0^U e^{-\frac{\alpha}{2}u}d\bar{B}_j(u)\right)^2\right] = \mathbb{E}\left[\int_0^U e^{-\alpha u}du\right] = \frac{1}{\alpha}(1 - e^{-\alpha U}), \tag{B.13}$$

which is bounded in U on $[0, +\infty)$. Thus, by L^2-bounded martingale convergence theorem,

$$\mathcal{I}_U \xrightarrow[U\to\infty]{a.s.} \mathcal{I}_\infty = \int_0^\infty e^{-\frac{\alpha}{2}u}d\bar{B}_j(u), \tag{B.14}$$

with $\mathbb{E}[\mathcal{I}_\infty^2] < \infty$. This can imply that $\mathbb{E}[A_j] \leqslant 2\left(\mathbb{E}[X_0] + \frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right) < \infty$. Then,

$$|Y_j(t) - \tilde{Y}_j(t)| \leqslant \sqrt{A_j}e^{-\frac{\alpha}{2}t} \text{ and } Y_j(t) - \tilde{Y}_j(t) \xrightarrow[t\to\infty]{a.s.} 0.$$

Then,

$$|Y_j^2(t) - \tilde{Y}_j^2(t)| \leqslant \sqrt{A_j}e^{-\frac{\alpha}{2}t}\left(|Y_j(t) + \tilde{Y}_j(t)|\right). \tag{B.15}$$

$$\sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_j^2(t) - \tilde{Y}_j^2(t)| \leqslant \sqrt{A_j} e^{-\frac{\alpha}{2} 2^n} \left(\sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_j(t)| + \sup_{2^n \leqslant t \leqslant 2^{n+1}} |\tilde{Y}_j(t)| \right).$$

Then, since the processes $\{X(t), t \geqslant 0\}$ and $\{\tilde{X}(t), t \geqslant 0\}$ have continuous trajectories,

$$\sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_j^2(t) - \tilde{Y}_j^2(t)| \leqslant \sqrt{A_j} e^{-\frac{\alpha}{2} 2^n} \left(|Y_j(t_{n1})| + |\tilde{Y}_j(t_{n2})| \right), \quad 2^n \leqslant t_{n1},\ t_{n2} \leqslant 2^{n+1}. \tag{B.16}$$

Further,

$$\sup_{t \geqslant 0} \mathbb{E}[|Y_j(t)|^2] \leqslant 2 \sup_{t \geqslant 0} \left(e^{-\alpha t} \mathbb{E}[|Y_j(0)|^2] + \frac{1}{2} \sigma e^{-\alpha t} \mathbb{E}\left[\left| \int_0^t e^{\frac{\alpha}{2} s} dB_j(s) \right|^2 \right] \right).$$

By Itô's isometry,

$$\mathbb{E}\left[\left| \int_0^t e^{\frac{\alpha}{2} s} dB_j(s) \right|^2 \right] = \mathbb{E}\left[\int_0^t e^{\alpha s} ds \right] = \frac{1}{\alpha}(e^{\alpha t} - 1), \tag{B.17}$$

which implies that $\sup_{t \geqslant 0} \mathbb{E}[|Y_j(t)|^2] < \infty$. Further,

$$\mathbb{E}\left[|\tilde{Y}_j(t)|^2 \right] \leqslant 2\mathbb{E}\left[\left| \frac{1}{2} \sigma e^{-\frac{\alpha}{2} t} \int_{-\infty}^0 e^{\frac{\alpha}{2} s} d\tilde{B}_j(s) \right|^2 \right] + 2\mathbb{E}\left[\left| \frac{1}{2} \sigma e^{-\frac{\alpha}{2} t} \int_0^t e^{\frac{\alpha}{2} s} d\tilde{B}_j(s) \right|^2 \right].$$

(B.12), (B.14) and (B.17) give $\sup_{t \geqslant 0} \mathbb{E}[|\tilde{Y}_j(t)|^2] < \infty$. Then,

$$\mathbb{E}\left[\left(|Y_j(t_{n1})| + |\tilde{Y}_j(t_{n2})| \right)^2 \right] \leqslant 2 \sup_{t \geqslant 0} \mathbb{E}[|Y_j(t)|^2] + 2 \sup_{t \geqslant 0} \mathbb{E}[|\tilde{Y}_j(t)|^2] \leqslant K_{1*} < \infty,$$

for some $K_{1*} > 0$. Finally, from (B.16) and Cauchy-Schwartz inequality,

$$\mathbb{E}\left[\sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_j^2(t) - \tilde{Y}_j^2(t)| \right] \leqslant \sqrt{2 \left(\mathbb{E}[X_0] + \frac{1}{4} \sigma^2 \mathbb{E}[\mathcal{I}_\infty^2] \right) K_{1*}} e^{-\frac{\alpha}{2} 2^n},$$

which implies that

$$\sum_{n=1}^{\infty} \mathbb{E}\left[\sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_j^2(t) - \tilde{Y}_j^2(t)| \right] \leqslant \sqrt{2 \left(\mathbb{E}[X_0] + \frac{1}{4} \sigma^2 \mathbb{E}[\mathcal{I}_\infty^2] \right) K_{1*}} \sum_{n=1}^{\infty} e^{-\frac{\alpha}{2} 2^n} < \infty.$$

From Markov inequality and Borel-Cantelli's lemma, $Y_j^2(t) - \tilde{Y}_j^2(t) \xrightarrow[t \to \infty]{a.s.} 0$. Hence,

$$\tilde{X}(t) - X(t) = \sum_{j=1}^d \tilde{Y}_j^2(t) - \sum_{j=1}^d Y_j^2(t) \xrightarrow[t \to \infty]{a.s.} 0.$$

(2). From (B.15), $\left|Y_j^2(t) - \tilde{Y}_j^2(t)\right| \leqslant \sqrt{A_j} e^{-\frac{\alpha}{2}t}\left(\left|Y_j(t) + \tilde{Y}_j(t)\right|\right)$. So,

$$\mathbb{E}\left[\left|\tilde{X}(t) - X(t)\right|^m\right] \leqslant \mathbb{E}\left[\left(\sum_{j=1}^d \left|Y_j^2(t) - \tilde{Y}_j^2(t)\right|\right)^m\right] \leqslant e^{-m\frac{\alpha}{2}t}\mathbb{E}\left[\left(\sum_{j=1}^d \sqrt{A_j}\left|Y_j(t) + \tilde{Y}_j(t)\right|\right)^m\right].$$

Then,

$$\mathbb{E}\left[\left|\tilde{X}(t) - X(t)\right|^m\right] \leqslant e^{-m\frac{\alpha}{2}t} d^{m-1}\left(\sum_{j=1}^d \mathbb{E}\left[\left(\sqrt{A_j}\left|Y_j(t) + \tilde{Y}_j(t)\right|\right)^m\right]\right).$$

By Cauchy-Schwartz inequality,

$$\mathbb{E}\left[\left(\sqrt{A_j}\left|Y_j(t) + \tilde{Y}_j(t)\right|\right)^m\right] \leqslant \sqrt{\mathbb{E}\left[A_j^m\right]\mathbb{E}\left[\left|Y_j(t) + \tilde{Y}_j(t)\right|^{2m}\right]}.$$

Further,

$$\begin{aligned}\mathbb{E}[A_j^m] &= 2^m\left(\mathbb{E}[Y_j^2(0)] + \tfrac{1}{4}\sigma^2\mathbb{E}\left[\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s} d\tilde{B}_j(s)\right)^2\right]\right)^m \\ &\leqslant 2^m 2^{m-1}\left(\mathbb{E}[X_0^m] + \left(\tfrac{1}{4}\sigma^2\right)^m \mathbb{E}\left[\mathcal{I}_\infty^{2m}\right]\right).\end{aligned}$$

Similar to (B.13), $\mathbb{E}\left[\mathcal{I}_U^{2m}\right] \leqslant \left(\frac{1}{\alpha}(1 - e^{-\alpha U})\right)^m$, which implies that $\mathbb{E}\left[\mathcal{I}_\infty^{2m}\right] < \left(\frac{1}{\alpha}\right)^m$. Then, there exists some positive constant K_{2*}, such that $\mathbb{E}[A_j^m] \leqslant K_{2*}$.

$$\left|Y_j(t) + \tilde{Y}_j(t)\right|^{2m} = \left|e^{-\frac{\alpha}{2}t} Y_j(0) + \sigma e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s} dB_j(s) + \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\int_{-\infty}^0 e^{\frac{\alpha}{2}s} d\tilde{B}_j(s)\right|^{2m}$$

gives

$$\begin{aligned}\mathbb{E}\left[\left|Y_j(t) + \tilde{Y}_j(t)\right|^{2m}\right] &\leqslant 3^{2m-1} e^{-\frac{\alpha}{2}2mt}\mathbb{E}\left[Y_j^{2m}(0) + \left|\sigma\int_0^t e^{\frac{\alpha}{2}s} dB_j(s)\right|^{2m}\right. \\ &\quad \left. + \left|\tfrac{1}{2}\sigma\int_{-\infty}^0 e^{\frac{\alpha}{2}s} d\tilde{B}_j(s)\right|^{2m}\right],\end{aligned}$$

with **(i)** $\mathbb{E}\left[Y_j^{2m}(0)\right] \leqslant \mathbb{E}\left[X_0^m\right] < \infty$; **(ii)** $\mathbb{E}\left[\left|\sigma\int_0^t e^{\frac{\alpha}{2}s} dB_j(s)\right|^{2m}\right] \leqslant \sigma^{2m} C_m\left(\frac{1}{\alpha}(e^{\alpha t} - 1)\right)^m$ for some $C_m > 0$; and **(iii)** $\mathbb{E}\left[\left|\frac{1}{2}\sigma\int_{-\infty}^0 e^{\frac{\alpha}{2}s} d\tilde{B}_j(s)\right|^{2m}\right] = \mathbb{E}[\mathcal{I}_\infty^{2m}] \leqslant \left(\frac{1}{\alpha}\right)^m$. Hence, there exists some positive constant K_{3^*}, such that $\mathbb{E}\left[\left|Y_j(t) + \tilde{Y}_j(t)\right|^{2m}\right] \leqslant K_{3^*}$. Then,

$$\mathbb{E}\left[|\tilde{X}(t) - X(t)|^m\right] \leqslant e^{-\frac{\alpha}{2}mt} d^{m/2}\sqrt{K_{2*}K_{3*}} \xrightarrow[t\to\infty]{} 0.$$

This finishes the proof of $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{L^m} 0$.

(3). It is obvious that both $\tilde{X}(t)$ and $X(t)$ are positive almost surely, for all $t \geq 0$.

Then, for all $t \geq 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(\tilde{X}(t) \leq \frac{1}{n}\right) = 0, \tag{B.18}$$

$$\lim_{n\to\infty} \mathrm{P}\left(X(t) \leq \frac{1}{n}\right) = 0. \tag{B.19}$$

Further,

$$\lim_{N\to\infty} \mathrm{P}\left(\tilde{X}(t)X(t) \leq \frac{1}{N}\right) = 0. \tag{B.20}$$

For $\forall\ \delta > 0$, let $A_\infty = \left\{\limsup_{t\to\infty} \frac{1}{\tilde{X}(t)X(t)} \left|\tilde{X}(t) - X(t)\right| > \delta\right\}$, and $A_N = \left\{\frac{1}{\tilde{X}(t)X(t)} > N\right\}$. $\mathrm{P}(A_\infty) = \mathrm{P}(A_\infty A_N) + \mathrm{P}(A_\infty A_N^c)$, where A_N^c is the complement of A_N. Then,

$$\mathrm{P}(A_\infty) \leq \mathrm{P}(A_N) + \mathrm{P}(A_\infty A_N^c) \leq \mathrm{P}(A_N) + \mathrm{P}\left(\limsup_{t\to\infty} \left|\tilde{X}(t) - X(t)\right| > \frac{1}{N}\delta\right).$$

Further, (B.20) implies that $\lim_{N\to\infty} \mathrm{P}(A_N) = 0$. From $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{a.s.} 0$, $\lim_{N\to\infty} \mathrm{P}\left(\limsup_{t\to\infty} \left|\tilde{X}(t) - X(t)\right| > \frac{1}{N}\delta\right) = 0$. Then, for $\forall\ \delta > 0$, $\mathrm{P}\left(\limsup_{t\to\infty} \left|\frac{1}{\tilde{X}(t)} - \frac{1}{X(t)}\right| > \delta\right) = \mathrm{P}(A_\infty) = 0$. This finishes the proof. □

B.2.2 General case

Lemma B.1. *Given a one-dimensional Brownian motion B defined on the probability space $(\Omega, \mathscr{F}, \mathrm{P})$ with $(\mathscr{F}_t)_{t\geqslant 0}$ as usual, there exists a two-dimensional Brownian motion $\tilde{W} = (\tilde{w}_1, \tilde{w}_2)$ on the same probability space, such that*

$$B_t = \sum_{j=1}^{2} \int_0^t \frac{X^{(i*)}(u)}{\|X^*(u)\|} d\tilde{w}_j(u) \quad \textit{almost surely.}$$

Furthermore, it is possible to choose $\tilde{W} = B^{(*)}$ almost surely.

Alternative Proof of Lemma B.1. Since for $i, j = 1, 2$, $\langle \tilde{w}_i, \tilde{w}_j\rangle_t = t\,\mathbb{I}_{\{i=j\}}$, Hence,

$$\left\langle \sum_{j=1}^{2} \int_0^{\cdot} \frac{X^{(j*)}(s)}{\sqrt{X^{(1*)}(s) + X^{(2*)}(s)}} d\tilde{w}_j(s), \sum_{j=1}^{2} \int_0^{\cdot} \frac{X^{(j*)}(s)}{\sqrt{X^{(1*)}(s) + X^{(2*)}(s)}} d\tilde{w}_j(s) \right\rangle_t = t.$$

It follows from the Lévy characterization for Brownian motion [Shreve, 2004, Theorem B.3] that $\sum_{j=1}^{2} \int_0^t \frac{X^{(j*)}(s)}{\sqrt{X^{(1*)}(s)+X^{(2*)}(s)}} d\tilde{w}_j(s)$ is a standard Brownian motion. This implies that

$$B_t = \sum_{j=1}^{2} \int_0^t \frac{X^{(j*)}(s)}{\sqrt{X^{(1*)}(s) + X^{(2*)}(s)}} d\tilde{w}_j(s), \quad \textit{almost surely.} \tag{B.21}$$

Furthermore, it is possible to choose $\tilde{W} = B^{(*)}$ almost surely. This completes the proof.

□

Proof of Proposition 3.2.4. From Lemma B.1, for the given Brownian motion B_t, the Brownian motion $B_t = \sum_{j=1}^{2} \int_0^t \frac{X^{(j*)}(s)}{\sqrt{X^{(1*)}(s)+X^{(2*)}(s)}} dB_s^{(j*)}$ and the process $X^{(1*)}(t) + X^{(2*)}(t)$ are constructed from it. Application of Itô's lemma to $X^{(1*)}(t) + X^{(2*)}(t)$ gives

$$\begin{aligned} d(X^{(1*)}(t) + X^{(2*)}(t)) &= \left(L(t) - \alpha \left(X^{(1*)}(t) + X^{(2*)}(t)\right)\right) dt \\ &+\sigma \left(\sqrt{X^{(1*)}(t)} dB^{(1*)}(t) + \sqrt{X^{(2*)}(t)} dB^{(2*)}(t)\right). \end{aligned}$$

From (B.21),

$$d(X^{(1*)}(t)) + X^{(2*)}(t)) = \left(L(t) - \alpha(X^{(1*)}(t) + X^{(2*)}(t)\right) dt + \sigma \sqrt{X^{(1*)}(t) + X^{(2*)}(t)} dB_t,$$

which shows that $X^{(1*)}(t) + X^{(2*)}(t)$ is a GCIR process. Then, $X^{(1*)}(t) + X^{(2*)}(t)$ is the strong solution of SDE (3.2.1). Because SDE (3.2.1) admits an unique strong solution, that is $X^{(1*)}(t) + X^{(2*)}(t) = X(t)$ almost surely. □

Proof of Proposition 3.2.5. The proof is similar to the proof of Lemma 2.2 in Tong and Zhang [2017]. Since the process in (3.2.8) is driven by Brownian motion, then, for fixed $t \in [0, T]$, $X^{(2*)}(t)(\omega)$ is a process with continuous sample paths, which implies that

$$\mathrm{P}\left(\omega : X^{(2*)}(t, w) = 0, t \geq 0\right) = 0.$$

By Fubini theorem,

$$\mathbb{E}\left[\int_0^T \mathbb{I}_{A_t}(w) dt\right] = \int_0^T \mathbb{E}\left[\mathbb{I}_{A_t}(w)\right] dt = \int_0^T \mathrm{P}\left(\omega : X^{(2*)}(t, w) = 0, t \geq 0\right) dt = 0.$$

Further, from $\int_0^T \mathbb{I}_{A_t}(w) dt \geq 0$, $\int_0^T \mathbb{I}_{A_t}(w) dt = 0$ a.s. □

Proof of Proposition 3.2.6. The solution to SDE (3.2.7) is clear from Section B.2.1. The dimension of (3.2.7) is 2. for $j = 1, 2$,

$$Y_j(t) = e^{-\frac{\alpha}{2}t} Y_j(0) + \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t} \int_0^t e^{\frac{\alpha}{2}s} dB_j(s).$$

Define $X^{(1*)}(t) = \sum_{j=1}^{2} Y_j^2(t)$, and notice that $B_t^{(1*)} = \sum_{j=1}^{2} \int_0^t \frac{Y_j(s)}{\sqrt{X^{(1*)}(t)}} dB_j(s)$. By using Itô's formula, one can prove that $X^{(1*)}(t)$ is the solution to the SDE (3.2.7), with $X_0^{(1*)} = \sum_{j=1}^{2} Y_j^2(0)$. For the solution to SDE (3.2.8), let $U(t, X^{(2*)}(t) = e^{\frac{\alpha}{2}t}\sqrt{X^{(2*)}(t)}$, by Itô's formula,

$$dU(t, X^{(2*)}(t) = \frac{\alpha}{2} e^{\frac{\alpha}{2}t} \sqrt{X^{(2*)}(t)} dt + e^{\frac{\alpha}{2}t} \frac{1}{2\sqrt{X^{(2*)}(t)}} dX^{(2*)}(t) - \frac{1}{8} e^{\frac{\alpha}{2}t} \frac{1}{\sqrt{\left(X^{(2*)}(t)\right)^{3/2}}} d\left\langle X^{(2*)}(t)\right\rangle$$

$$= \frac{\alpha}{2} e^{\frac{\alpha}{2}t} \sqrt{X^{(2*)}(t)} dt + e^{\frac{\alpha}{2}t} \frac{1}{2\sqrt{X^{(2*)}(t)}} \left(\left(L(t) - \frac{\sigma^2}{2} - \alpha X^{(2*)}(t) \right) dt + \sigma \sqrt{X^{(2*)}(t)} dB_t^{(2*)} \right)$$

$$-\frac{1}{8} e^{\frac{\alpha}{2}t} \frac{1}{\sqrt{\left(X^{(2*)}(t)\right)^{3/2}}} \sigma^2 X^{(2*)}(t) dt.$$

Then,

$$dU(t, X^{(2*)}(t) = e^{\frac{\alpha}{2}t} \frac{1}{2\sqrt{X^{(2*)}(t)}} \left(L(t) - \frac{\sigma^2}{2} \right) dt$$

$$+ e^{\frac{\alpha}{2}t} \frac{1}{2\sqrt{X^{(2*)}(t)}} \sigma \sqrt{X^{(2*)}(t)} dB_t^{(2*)} - \frac{\sigma^2}{8} e^{\frac{\alpha}{2}t} \frac{1}{\sqrt{X^{(2*)}(t)}} dt.$$

This gives

$$d\left(e^{\frac{\alpha}{2}t} \sqrt{X^{(2*)}(t)}\right) = e^{\frac{\alpha}{2}t} \frac{1}{2\sqrt{X^{(2*)}(t)}} \left(L(t) - \frac{3\sigma^2}{4} \right) dt + \frac{\sigma}{2} e^{\frac{\alpha}{2}t} dB_t^{(2*)}.$$

Integrating both sides from 0 to t and multiplying by $e^{-\frac{\alpha}{2}t}$,

$$Y_t = \sqrt{X^{(2*)}(t)} = e^{-\frac{\alpha}{2}t} \sqrt{X_0^{(2*)}} + e^{-\frac{\alpha}{2}t} \int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}} ds + \frac{\sigma}{2} e^{-\frac{\alpha}{2}t} \int_0^t e^{\frac{\alpha}{2}s} dB_s^{(2*)}.$$

Then, the solution has the following expression

$$X^{(2*)}(t) = Y_t^2 = \left(e^{-\frac{\alpha}{2}t} \sqrt{X_0^{(2*)}} + e^{-\frac{\alpha}{2}t} \int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}} ds + \frac{\sigma}{2} e^{-\frac{\alpha}{2}t} \int_0^t e^{\frac{\alpha}{2}s} dB_s^{(2*)} \right)^2.$$

□

Proof of Proposition 3.3.3. It is clear that $\tilde{Y}_1(t)$ and $\tilde{Y}_2(t)$ are independent with $\tilde{Y}_j(t) \sim \mathcal{N}\left(0, \frac{\sigma^2}{4\alpha}\right)$, $j = 1, 2$. This implies that $\tilde{X}^{(1*)}(t) = \frac{\sigma^2}{4\alpha} \sum_{j=1}^{2} \frac{4\alpha}{\sigma^2} \tilde{Y}_j^2(t) \sim \frac{\sigma^2}{4\alpha}\chi_2^2$ and then $\{\tilde{X}^{(1*)}(t), t \geqslant 0\}$ is strictly stationary and ergodic. Below, by using the extension of Dubins-Schwarz Theorem, one proves that $\{\tilde{Y}_{t+k-1}, t \in (0,1]\}_{k\in\mathbb{N}_+}$ is strictly stationary and ergodic, which implies that $\{\tilde{X}^{(2*)}(t+k-1), t \in (0,1]\}_{k\in\mathbb{N}_+}$is strictly stationary and ergodic. To this end, note that, from (3.3.3),

$$e^{-\frac{\alpha}{2}t} \int_{-\infty}^{t} e^{\frac{\alpha}{2}s} \mathcal{V}(s) ds = e^{-\frac{\alpha}{2}t} \left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8} ds + \int_{0}^{t} e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}} ds \right)$$

and note that $\int_0^t e^{\frac{\alpha}{2}s}(4L(s) - 3\sigma^2)/\left(8\sqrt{X^{(2*)}(s)}\right) ds$ is the quadratic variation of the process
$\int_0^t e^{\frac{\alpha}{4}s} \sqrt{(4L(s) - 3\sigma^2)/\left(8\sqrt{X^{(2*)}(s)}\right)} dB_s$, where B_s is the Brownian motion, i.e.

$$\int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}} ds = \left\langle \int_0^{\cdot} e^{\frac{\alpha}{4}s} \sqrt{\frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}}} dB_s \right\rangle_t, \quad a.s.$$

Let $M(t) = \int_0^t e^{\frac{\alpha}{4}s} \sqrt{\frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}}} dB_s$. Since Itô integrals are local martingales [Klebaner, 2005, Theorem 13.2], $M(t)$ is a local martingale. with $\langle M\rangle_t = \int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s)-3\sigma^2}{8\sqrt{X^{(2*)}(s)}} ds$, and

$$\mathbb{E}\left[\langle M\rangle_t\right] = \mathbb{E}\left[\int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(s)}} ds\right] \leqslant \int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8} \mathbb{E}\left[\frac{1}{\sqrt{X^{(2*)}(s)}}\right] ds.$$

From (B.32), $\mathbb{E}\left[\frac{1}{\sqrt{X^{(2*)}(s)}}\right] \leqslant \sqrt{\sup_{t\geqslant 0} \mathbb{E}\left[\frac{1}{X^{(2*)}(s)}\right]} < \infty$, which implies that $\mathbb{E}\left[|M(t)|\right] < \infty$. Thus, $M(t)$ is a continuous martingale and $M_0 = 0$, $\langle M\rangle_\infty = \infty$ almost surely. For all $t > 0$, let $\phi(t) = \int_0^t e^{\frac{\alpha}{2}s} \frac{4L(s)-3\sigma^2}{8} ds$. It is a homeomorphism on $0 \leqslant t < \infty$ and $\lim_{t\to\infty} \phi(t) = \infty$ and, $\phi^{-1}(t)$ is continuous. Consider the stopping time $\tau(t) = \inf\{s > 0 : \langle M\rangle_s > \phi(t)\}$. Then, by extension of Dubins-Schwarz Theorem, the time changed process $N(t) := M(\tau(t)), 0 \leqslant t < \infty$ is a $(\Omega, \mathcal{G}_t, \mathrm{P})$ square-integrable martingale with $\mathcal{G}_t = \mathcal{F}_{\tau(t)}$ and can be represented as $N(t) = \int_0^t e^{\frac{\alpha}{4}s} \sqrt{(4L(s) - 3\sigma^2)/8}\, d\boldsymbol{B}_s$ a.s, where $\boldsymbol{B}_t$ is a $\mathcal{G}_t$ measurable

Brownian motion. $\langle M\rangle_{\tau(t)} = \phi(t) = \langle N\rangle_t$. $N(t) = M(\tau(t))$ implies that $N(\langle M\rangle_t) = M(\tau(\langle M\rangle_t))$. Let $f(t) = \langle M\rangle_t$ be the quadratic variation of $M(t)$, f is an invertible and increasing function. Further, $\tau(t) = f^{-1}(\phi(t))$ with probability 1. This implies that $\tau^{-1}(t) = \phi^{-1}(f(t))$ with probability 1. Then,

$$\tau\left(\tau^{-1}(t)\right) = f^{-1}\left(\phi(\tau^{-1}(t))\right) = f^{-1}\left(\phi(\phi^{-1}(f(t)))\right) = f^{-1}(f(t)) = t.$$

Then

$$M(t) = N(\tau^{-1}(t)), \quad a.s.$$

Then, for $t \in (0, 1], k \in \mathbb{N}_+$,

$$M(t+k-1) = \int_0^{t+k-1} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8\sqrt{X^{(2*)}(s)}}}dB_s = \int_0^{\tau^{-1}(t+k-1)} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\boldsymbol{B}_s,$$

and then

$$M(t+k-1) = \int_0^{\tau^{-1}(k-1)} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\boldsymbol{B}_s + \int_{\tau^{-1}(k-1)}^{\tau^{-1}(t+k-1)} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\boldsymbol{B}_s.$$

Further,

$$\int_{-\infty}^{t+k-1} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds = \int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds + \int_0^{t+k-1} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds,$$

and $\int_0^{t+k-1} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds = \left\langle \int_0^{\cdot} e^{\frac{\alpha}{4}s}\sqrt{\mathcal{V}(s)}dB_s\right\rangle_{t+k-1}$. Let $u = -s$,

$$\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds = \int_0^{\infty} e^{-\frac{\alpha}{2}u}\mathcal{V}(-u)du.$$

Let $\mathbb{Y}(U) = \displaystyle\int_0^U e^{-\frac{\alpha}{4}u}\sqrt{\frac{4L(s)-3\sigma^2}{8}}dB_u$. Then, $\mathbb{Y}(U)$ is a continuous local martingale with $\mathbb{E}[|\mathbb{Y}(0)|^2] < \infty$. By Doob-Meyer decomposition, $\mathbb{Y}^2(U) - \langle\mathbb{Y}\rangle_U$ is an uniformly integrable martingale, and $\mathbb{E}\left[\mathbb{Y}^2(\infty) - \langle\mathbb{Y}\rangle_\infty\right] = 0$, which implies that $\mathbb{E}\left[\langle\mathbb{Y}\rangle_\infty\right] = \mathbb{E}\left[\mathbb{Y}^2(\infty)\right]$. Since $\langle\mathbb{Y}\rangle_\infty$ is not random, $\mathbb{E}\left[\langle\mathbb{Y}\rangle_\infty\right] = \langle\mathbb{Y}\rangle_\infty$. This gives $\langle\mathbb{Y}\rangle_\infty = \mathbb{E}\left[\mathbb{Y}^2(\infty)\right] = \displaystyle\int_0^\infty e^{-\frac{\alpha}{2}u}\frac{4L(-s)-3\sigma^2}{8}du$. Let $s = -u$,

$$\int_0^\infty e^{-\frac{\alpha}{4}u}\sqrt{\frac{4L(-s)-3\sigma^2}{8}}dB_u = \int_{-\infty}^0 e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\bar{B}_{-s}.$$

Then,

$$e^{-\frac{\alpha}{2}(t+k-1)}\int_{-\infty}^{t+k-1} e^{\frac{\alpha}{2}s}\mathcal{V}(s)ds = e^{-\frac{\alpha}{2}(t+k-1)}\left\langle \int_{-\infty}^{\cdot} e^{\frac{\alpha}{4}s}\sqrt{\mathcal{V}(s)}d\tilde{B}_s\right\rangle_{t+k-1}.$$

Further,

$$\int_{-\infty}^{t+k-1} e^{\frac{\alpha}{4}s}\sqrt{\mathcal{V}(s)}d\tilde{B}_s = \int_{-\infty}^{0} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\tilde{B}_s + \int_{0}^{t+k-1} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8\sqrt{X^{(2*)}(t)}}}dB_s$$

$$= \int_{-\infty}^{0} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\tilde{B}_s + \int_{0}^{\tau^{-1}(t+k-1)} e^{\frac{\alpha}{4}s}\sqrt{\frac{4L(s)-3\sigma^2}{8}}d\boldsymbol{B}_s.$$

This gives

$$\tilde{Y}_{t+k-1} = e^{-\frac{\alpha}{2}(t+k-1)}\int_{-\infty}^{\tau^{-1}(t+k-1)} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds + \frac{\sigma}{2}e^{-\frac{\alpha}{2}t+k-1}\int_{-\infty}^{t+k-1} e^{\frac{\alpha}{2}s}d\tilde{B}_s$$

$$= e^{-\frac{\alpha}{2}t}\int_{-\infty}^{\tau^{-1}(t)} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds + \frac{\sigma}{2}e^{-\frac{\alpha}{2}(t+k-1)}\int_{-\infty}^{t+k-1} e^{\frac{\alpha}{2}s}d\tilde{B}_s.$$

Further,

$$e^{-\frac{\alpha}{2}(t+k-1)}\int_{-\infty}^{\tau^{-1}(t+k-1)} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds = e^{-\frac{\alpha}{2}t}\int_{-\infty}^{\tau^{-1}(t)} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds$$

implies that the random sequence $\left\{e^{-\frac{\alpha}{2}(t+k-1)}\int_{-\infty}^{\tau^{-1}(t+k-1)} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds, t\in(0,1]\right\}_{k\in\mathbb{N}_+}$ is stationary in strict sense. For fixed t, $e^{-\frac{\alpha}{2}t}\int_{-\infty}^{\tau^{-1}(t)} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds$ is a random variable, which is independent of k. In addition, $\left\{\frac{\sigma}{2}e^{-\frac{\alpha}{2}(t+k-1)}\int_{-\infty}^{t+k-1} e^{\frac{\alpha}{2}s}d\tilde{B}_s, t\in[0,1]\right\}_{k\in\mathbb{N}_+}$ is strictly stationary and ergodic. Hence $\tilde{Y}_{t+k-1}$ is strictly stationary and ergodic. This implies that $\left\{\tilde{X}^{(2*)}(t+k-1), t\in(0,1]\right\}_{k\in\mathbb{N}_+}$ is strictly stationary and ergodic and since $\left\{\tilde{X}^{(1*)}(t+k-1), t\in(0,1]\right\}_{k\in\mathbb{N}_+}$ is also strictly stationary and ergodic, it concludes that $\left\{\tilde{X}(t+k-1) = \tilde{X}^{(1*)}(t+k-1) + \tilde{X}^{(2*)}(t+k-1), 0 < t \leqslant 1\right\}_{k\in\mathbb{N}_+}$ is strictly stationary and ergodic. □

B.3 Point estimators and their asymptotic properties

B.3.1 On derivation of UMLE

Proof of Proposition 3.3.1. Let $a = [a_1^\top, a_2]$ with a_1 a $p-$column vector, and a_2 a scalar. Then,

$$aQ_{[0,T]}a^\top = \left(a_1^\top \int_0^T \frac{\varphi^\top(t)\varphi(t)}{X(t)}dt - a_2 \int_0^T \varphi(t)dt, -a_1^\top \int_0^T \varphi^\top(t)dt + a_2 \int_0^T X(t)dt\right)\times (a_1^\top, a_2)^\top.$$

Then, $aQ_{[0,T]}a^\top = \int_0^T \left(a_1^\top \frac{\varphi^\top(t)}{\sqrt{X(t)}} - a_2\sqrt{X(t)}\right)^2 dt \geqslant 0$. The equality hold if and only if $\left(a_1^\top \frac{\varphi^\top(t)}{\sqrt{X(t)}} - a_2\sqrt{X(t)}\right)^2 = 0$, almost everywhere (a.e) for $t \in [0, T]$, which is $a_1^\top \frac{\varphi^\top(t)}{\sqrt{X(t)}} - a_2\sqrt{X(t)} = 0$, a.e for $t \in [0, T]$ and then, $\mathrm{P}\left(\omega : a_1^\top \frac{\varphi^\top(t)}{\sqrt{X(t,\omega)}} - a_2\sqrt{X(t,\omega)} = 0, \forall\, t \in [0, T]\right) = 1$, which is equivalent to the following

$$\mathrm{P}\left(\omega : a_1^\top\varphi^\top(t) - a_2X(t,\omega) = 0, \forall\, t \in [0, T]\right) = 1, \tag{B.22}$$

and since the process $\{X(t), t \geqslant 0\}$ has absolutely continuous distribution, for $a_2 \neq 0$,

$$\mathrm{P}\left(\omega : a_1^\top\varphi^\top(t) - a_2X(t,\omega) = 0, \forall\, t \in [0, T]\right) = 0.$$

This is a contradiction with (B.22), hence, $a_2 = 0$. Therefore, putting $a_2 = 0$ into $a_1^\top \frac{\varphi^\top(t)}{\sqrt{X(t)}} - a_2\sqrt{X(t)} = 0$, provided that $T \geqslant 1$, that $a_1^\top \frac{\varphi^\top(t)}{\sqrt{X(t)}} = 0$, which is equivalent to $a_1^\top\varphi^\top(t) = 0$. By Assumption 2.2, the base function $\varphi(t)$ is linearly independent, which means $a_1^\top\varphi^\top(t) = 0$ implies that $a_1 = \vec{0}_{(p_0+1)\times 1}$. This completes the proof. □

Proof of the Proposition 3.3.2. The likelihood function of the SDE (3.2.1) is given by

$$\mathcal{L}(\theta, X^T) = \frac{d\mathrm{P}^{(\theta)}_{X^T}}{dP_B} = \exp\left(\frac{1}{\sigma^2}\int_0^T \frac{S(t,\theta,X(t))}{X(t)}dX(t) - \frac{1}{2\sigma^2}\int_0^T \frac{S^2(t,\theta,X(t))}{X(t)}dt\right).$$

Together with some algebraic computations, the log-likelihood function is

$$\log\mathcal{L}(\theta, X^T) = \frac{1}{\sigma^2}\theta^\top R_{[0,T]} - \frac{1}{2\sigma^2}\theta^\top Q_{[0,T]}\theta.$$

Then, by combining Proposition 3.3.1 and some algebraic computations, the UMLE is derived as $\hat{\theta}_T = Q^{-1}_{[0,T]} R_{[0,T]}$.

For the deriving of RMLE, it is necessary to find the maximum value of the log likelihood function Further, by Lagrange multiplier method, one maximizes $\log \mathcal{L}(\theta, X^T)$ under the restriction (2.2.2), which yield the RMLE as stated. This completes the proof. □

B.3.2 Proofs related to asymptotic distribution of estimators

Proof of Proposition 3.3.4. From Proposition B.3,

$$\tilde{X}^{(1*)}(t) - X^{(1*)}(t) \xrightarrow[t\to\infty]{a.s.\ and\ L^m} 0. \tag{B.23}$$

Next step is to derive the convergence of $X^{(2*)}(t) - \tilde{X}^{(2*)}(t) = Y_t^2 - \tilde{Y}_t^2$.

$$|Y_t - \tilde{Y}_t|^2 = \left(e^{-\frac{\alpha}{2}t}\sqrt{X_0^{(2*)}} + e^{-\frac{\alpha}{2}t}\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds + \frac{1}{2}\sigma e^{-\frac{\alpha}{2}t}\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} d\tilde{B}_s^{(2*)}\right)^2$$

$$\leqslant e^{-\alpha t}\left(\sqrt{X_0} + \int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds + \frac{1}{2}\sigma\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} d\tilde{B}_s^{(2*)}\right)^2$$

$$\leqslant 3e^{-\alpha t}\left(X_0 + \left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds\right)^2 + \frac{1}{4}\sigma^2\left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} d\tilde{B}_s^{(2*)}\right)^2\right).$$

Then $|Y_t - \tilde{Y}_t| \leqslant \sqrt{\mathbb{A}}e^{-\frac{\alpha}{2}t}$ with $\mathbb{A} = 3\left(X_0 + \left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds\right)^2 + \frac{1}{4}\sigma^2\left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} d\tilde{B}_s^{(2*)}\right)^2\right)$.

$$\mathbb{E}[\mathbb{A}] = 3\left(\mathbb{E}[X_0] + \left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8}ds\right)^2 + \frac{1}{4}\sigma^2\mathbb{E}\left[\left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} d\tilde{B}_s^{(2*)}\right)^2\right]\right).$$

$\forall s < 0$, $\tilde{B}_s^{(2*)} = \bar{B}_{-s}^{(2*)}$,

$$\mathbb{E}\left[\left(\int_{-\infty}^{0} e^{\frac{\alpha}{2}s} d\bar{B}_s^{(2*)}\right)^2\right] = \mathbb{E}\left[\left(\int_{0}^{\infty} e^{-\frac{\alpha}{2}u} d\bar{B}_u^{(2*)}\right)^2\right]. \tag{B.24}$$

Let $\mathcal{I}_U = \int_0^U e^{-\frac{\alpha}{2}u} d\bar{B}_u^{(2*)}$. By Itô's isometry,

$$\mathbb{E}[\mathcal{I}_U^2] = \mathbb{E}\left[\left(\int_0^U e^{-\frac{\alpha}{2}u} d\bar{B}_u^{(2*)}\right)^2\right] = \mathbb{E}\left[\int_0^U e^{-\alpha u}du\right] = \frac{1}{\alpha}(1 - e^{-\alpha U}), \tag{B.25}$$

which is bounded in U on $[0, +\infty)$. Thus, by L^2-bounded martingale convergence theorem,

$$\mathcal{I}_U \xrightarrow[U\to\infty]{a.s.} \mathcal{I}_\infty = \int_0^\infty e^{-\frac{\alpha}{2}u} d\bar{B}_u^{(2*)}, \tag{B.26}$$

with $\mathbb{E}[\mathcal{I}_\infty^2] < \infty$. In addition, by Assumption 2.2,

$$\left(\int_{-\infty}^0 e^{\frac{\alpha}{2}s} \frac{4L(s) - 3\sigma^2}{8} ds\right)^2 \leqslant \left(\frac{4K_\mu K_\varphi - 3\sigma^2}{4\alpha}\right)^2.$$

This implies that $\mathbb{E}[A] \leqslant 3\left(\mathbb{E}[X_0] + \left(\frac{4K_\mu K_\varphi - 3\sigma^2}{8}\frac{2}{\alpha}\right)^2 + \frac{1}{4}\sigma^2 \mathbb{E}[\mathcal{I}_\infty^2]\right) < \infty$. Then,

$$|Y_t - \tilde{Y}_t| \leqslant \sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}t} \text{ and } Y_t - \tilde{Y}_t \xrightarrow[t\to\infty]{a.s.} 0.$$

Further,

$$|Y_t^2 - \tilde{Y}_t^2| \leqslant \sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}t}\left(|Y_t + \tilde{Y}_t|\right) \leqslant \sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}t}\left(\sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}t} + 2|Y_t|\right) = \mathbb{A} e^{-\alpha t} + 2\sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}t}|Y_t|, \tag{B.27}$$

by using the fact that $|Y_t + \tilde{Y}_t| = |Y_t + \tilde{Y}_t - Y_t + Y_t| \leqslant \sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}t} + 2|Y_t|$.

$$\sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_t^2 - \tilde{Y}_t^2| \leqslant \mathbb{A} e^{-\alpha 2^n} + \sqrt{\mathbb{A}} e^{-\frac{\alpha}{2}2^n}\left(2 \sup_{2^n \leqslant t \leqslant 2^{n+1}} |Y_t|\right).$$

In addition, $Y_t = e^{-\frac{\alpha}{2}t}\sqrt{X_0^{(2*)}} + e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}\frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(t)}} ds + \frac{\sigma}{2} e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s} dB_s^{(2*)}$ implies that

$$\begin{aligned} |Y_t| &\leqslant \left|e^{-\frac{\alpha}{2}t}\sqrt{X_0^{(2*)}}\right| + \left|e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}\frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(t)}} ds\right| + \left|\frac{\sigma}{2} e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s} dB_s^{(2*)}\right| \\ &\leqslant \left|e^{-\frac{\alpha}{2}t}\sqrt{X_0}\right| + \left|e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}\frac{4L(s) - 3\sigma^2}{8\sqrt{X^{(2*)}(t)}} ds\right| + \left|\frac{\sigma}{2} e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s} dB_s^{(2*)}\right|. \end{aligned}$$

Let $\mathcal{Y}_t = \left|e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s}\frac{4L(s)-3\sigma^2}{8\sqrt{X^{(2*)}(t)}} ds\right| + \left|\frac{\sigma}{2} e^{-\frac{\alpha}{2}t}\int_0^t e^{\frac{\alpha}{2}s} dB_s^{(2*)}\right|$. Since the processes $\{X(t), t \geqslant 0\}$ and $\{\tilde{X}(t), t \geqslant 0\}$ have continuous trajectories,

$$\sup_{2^n \leqslant t \leqslant 2^{n+1}} |\mathcal{Y}_t| \leqslant |\mathcal{Y}_{t_{n1}}|, \quad 2^n \leqslant t_{n1} \leqslant 2^{n+1}. \tag{B.28}$$

Further,

$$\sup_{t\geqslant 0}\mathbb{E}[|\mathcal{Y}_t|] \leqslant \sup_{t\geqslant 0}\Bigg(e^{-\frac{\alpha}{2}t}\mathbb{E}\left[\int_0^t e^{\frac{\alpha}{2}s}(4L(s)-3\sigma^2)\Big/(8\sqrt{X^{(2*)}(t)})ds\right] + \frac{\sigma}{2}e^{-\frac{\alpha}{2}t}\mathbb{E}\left[\left|\int_0^t e^{\frac{\alpha}{2}s}dB_s^{(2*)}\right|\right]\Bigg),$$

and

$$e^{-\frac{\alpha}{2}t}\mathbb{E}\left[\int_0^t \frac{e^{\frac{\alpha}{2}s}(4L(s)-3\sigma^2)}{8\sqrt{X^{(2*)}(t)}}ds\right] = e^{-\frac{\alpha}{2}t}\int_0^t \frac{e^{\frac{\alpha}{2}s}(4L(s)-3\sigma^2)}{8}\mathbb{E}\left[1\Big/\sqrt{X^{(2*)}(t)}\right]ds.$$

From Assumption 3.2, the dimension of model (3.2.8) is greater than 2. (B.32) implies that

$$\mathbb{E}\left[1\Big/\sqrt{X^{(2*)}(t)}\right] \leqslant \sqrt{\mathbb{E}\left[1\Big/X^{(2*)}(t)\right]} \leqslant \sqrt{\sup_{s\geqslant 0}\mathbb{E}\left[1\Big/X^{(2*)}(s)\right]} < \infty.$$

By Itô's isometry,

$$\mathbb{E}\left[\left|\int_0^t e^{\frac{\alpha}{2}s}dB_s^{(2*)}\right|^2\right] = \mathbb{E}\left[\int_0^t e^{\alpha s}ds\right] = \frac{1}{\alpha}(e^{\alpha t}-1). \tag{B.29}$$

Hence, by combining (B.12), (B.14) and (B.29), $\sup_{t\geqslant 0}\mathbb{E}[|\mathcal{Y}_t|] \leqslant \mathbb{K}_{1*} < \infty$. Then,

$$\mathbb{E}\left[\sup_{2^n\leqslant t\leqslant 2^{n+1}}|\mathcal{Y}_t|\right] \leqslant \mathbb{E}\left[|\mathcal{Y}_{t_{n1}}|\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}[|\mathcal{Y}_t|] \leqslant \mathbb{K}_{1*} < \infty,$$

and further,

$$\mathbb{A}e^{-\alpha t} + 2\sqrt{\mathbb{A}}e^{-\frac{\alpha}{2}t}|Y_t| \leqslant \mathbb{A}e^{-\alpha t} + 2\sqrt{\mathbb{A}}e^{-\frac{\alpha}{2}t}\left(\left|e^{-\frac{\alpha}{2}t}\sqrt{X_0}\right| + |\mathcal{Y}_t|\right).$$

Because $\mathbb{A}$ and $\mathcal{Y}_t$ are independent, from (B.27) and Cauchy-Schwartz inequality,

$$\begin{aligned}\mathbb{E}\left[\sup_{2^n\leqslant t\leqslant 2^{n+1}}|Y^2(t)-\tilde{Y}^2(t)|\right] &\leqslant 3\left(\mathbb{E}[X_0] + \left(\frac{4K_\mu K_\varphi - 3\sigma^2}{8}\frac{2}{\alpha}\right)^2 + \frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)e^{-\alpha 2^n} \\ &+2\sqrt{3\left(\mathbb{E}[X_0^2] + \left(\frac{4K_\mu K_\varphi - 3\sigma^2}{8}\frac{2}{\alpha}\right)^2\mathbb{E}[X_0] + \frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\mathbb{E}[X_0]\right)e^{-\alpha 2^n}} \\ &+\sqrt{3\left(\mathbb{E}[X_0] + \left(\frac{4K_\mu K_\varphi - 3\sigma^2}{8}\frac{2}{\alpha}\right)^2 + \frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)}e^{-\frac{\alpha}{2}2^n}\left(2\mathbb{K}_{1*}\right),\end{aligned}$$

by using $\mathbb{E}\left[\sqrt{\mathbb{A}}\right] \leqslant \sqrt{\mathbb{E}\left[\mathbb{A}\right]}$ and $\mathbb{E}\left[\sqrt{\mathbb{A}X_0}\right] \leqslant \sqrt{\mathbb{E}\left[\mathbb{A}X_0\right]}$. This implies that

$$\sum_{n=1}^{\infty}\mathbb{E}\left[\sup_{2^n\leqslant t\leqslant 2^{n+1}}|Y_j^2(t)-\tilde{Y}_j^2(t)|\right] \leqslant 3\left(\mathbb{E}[X_0]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)\sum_{n=1}^{\infty}e^{-\alpha 2^n}$$
$$+2\sqrt{3\left(\mathbb{E}[X_0^2]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2\mathbb{E}[X_0]+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\mathbb{E}[X_0]\right)}\sum_{n=1}^{\infty}e^{-\alpha 2^n}$$
$$+\sqrt{3\left(\mathbb{E}[X_0]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)}\sum_{n=1}^{\infty}e^{-\frac{\alpha}{2}2^n}<\infty.$$

From Borel-Cantelli's lemma, $Y_t^2-\tilde{Y}_t^2 \xrightarrow[t\to\infty]{a.s.} 0$. Hence,

$$X^{(2*)}(t)-\tilde{X}^{(2*)}(t)=Y_t^2-\tilde{Y}_t^2 \xrightarrow[t\to\infty]{a.s.} 0. \tag{B.30}$$

(2) From (B.27),

$$|\tilde{X}^{(2*)}(t)-X^{(2*)}(t)|=|Y_t^2-\tilde{Y}_t^2|\leqslant \mathbb{A}e^{-\alpha t}+2\sqrt{\mathbb{A}}e^{-\frac{\alpha}{2}t}|Y_t|.$$

From $\mathbb{E}[\mathbb{A}]\leqslant 3\left(\mathbb{E}[X_0]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)<\infty$ and $\sup_{t\geqslant 0}\mathbb{E}[|Y_t^r|]\leqslant \mathbb{K}_{1*}<\infty$,

$$\mathbb{E}\left[|\tilde{X}^{(2*)}(t)-X^{(2*)}(t)|\right]\leqslant 3\left(\mathbb{E}[X_0]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)e^{-\alpha t}$$
$$+2\sqrt{3\left(\mathbb{E}[X_0^2]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2\mathbb{E}[X_0]+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\mathbb{E}[X_0]\right)}e^{-\alpha t}$$
$$+2\sqrt{3\left(\mathbb{E}[X_0]+\left(\frac{4K_\mu K_\varphi-3\sigma^2}{8}\frac{2}{\alpha}\right)^2+\frac{1}{4}\sigma^2\mathbb{E}[\mathcal{I}_\infty^2]\right)}e^{-\frac{\alpha}{2}t}\mathbb{K}_{1*}\xrightarrow[t\to\infty]{}0.$$

Then,

$$\tilde{X}^{(2*)}(t)-X^{(2*)}(t)\xrightarrow[t\to\infty]{L^1}0. \tag{B.31}$$

Then, (B.23), (B.30) and (B.31) imply that

$$\tilde{X}(t)-X(t)=\left(\tilde{X}^{(1*)}(t)-X^{(1*)}(t)\right)+\left(\tilde{X}^{(2*)}(t)-X^{(2*)}(t)\right)\xrightarrow[t\to\infty]{a.s.\ and\ L^1}0.$$

This completes the proof. □

Proof of Proposition 3.3.5. Since

$$\left|\tilde{X}(t)\right| = \left|\tilde{X}(t) - X(t) + X(t)\right| \leq \left(\left|\tilde{X}(t) - X(t)\right| + |X(t)|\right).$$

From Proposition 3.3.4,

$$\sup_{t\geq 0} \mathbb{E}\left[\left|\tilde{X}(t) - X(t)\right|\right] \leq \sup_{t\geq 0}\left(\mathbb{E}\left[\left|\tilde{X}^{(1*)}(t) - X^{(1*)}(t)\right|\right] + \mathbb{E}\left[\left|\tilde{X}^{(2*)}(t) - X^{(2*)}(t)\right|\right]\right) < \infty.$$

By Proposition 3.2.3, $\sup_{t\geq 0} \mathbb{E}[X^m(t)] \leq K_m$. Then,

$$\sup_{t\geq 0} \mathbb{E}\left[\left|\tilde{X}(t)\right|\right] \leq \left(\sup_{t\geq 0} \mathbb{E}\left[\left|\tilde{X}(t) - X(t)\right|\right] + \sup_{t\geq 0} \mathbb{E}\left[|X(t)|\right]\right) < \infty.$$

This completes the proof. □

Proof of Proposition 3.3.6. $\tilde{X}_0 = \tilde{X}_0^{(1*)} + \tilde{X}_0^{(2*)} \sim \frac{\sigma^2}{4\alpha}\chi^2(3, \Delta^*)$. This implies that $\mathbb{E}\left[1/\tilde{X}_0\right] < 4\alpha/\sigma^2 < \infty$. Further, for $0 < t \leqslant 1$, by Fatou's Lemma, that

$$\mathbb{E}\left[1/\tilde{X}(t)\right] = \mathbb{E}\left[\lim_{k\to\infty} 1/X(t+k-1)\right] \leqslant \liminf_{k} \mathbb{E}\left[1/X(t+k-1)\right] \leqslant \sup_{t\geqslant 0} \mathbb{E}\left[1/X(t)\right].$$

By Proposition 3 in Alaya and Kebaier [2013], $\sup_{t\geqslant 0} \mathbb{E}\left[(X(t))^\eta\right] < \infty$ for $\eta \in \left[-\frac{2\alpha\beta}{\sigma^2}, +\infty\right)$. By Proposition B.2 and Assumption 3.2, that $X(t) \leqslant X(t)$ a.s. which implies that

$$\sup_{t\geqslant 0} \mathbb{E}\left[(X(t))^\eta\right] \leqslant \sup_{t\geqslant 0} \mathbb{E}\left[(X(t))^\eta\right] < \infty, \quad \text{for } \eta \in \left[-\frac{2\alpha\beta}{\sigma^2}, 0\right), \tag{B.32}$$

which implies that $\sup_{t\geqslant 0} \mathbb{E}\left[1/X(t)\right] < \infty$. Then, $\sup_{t\geqslant 0} \mathbb{E}\left[1/\tilde{X}(t)\right] < \infty$. □

Proof of Proposition 3.3.8. Let $a = [a_1^\top, a_2]$ with a_1 a $p-$column vector, and a_2 a scalar. The matrix Σ is a positive definite matrix, as long as $a\Sigma a^\top > 0$ for any vector a. Then, from the definition of a and Σ,

$$a\Sigma a^\top = a_1 \int_0^1 \varphi^\top(t)\varphi(t)\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right] dt a_1 - 2a_2 \int_0^1 \varphi(t) dt a_1 + a_2 \int_0^1 \mathbb{E}\left[\tilde{X}(t)\right] dt a_2.$$

Then,

$$a\Sigma a^\top = \int_0^1 \left(a_1\varphi^\top(t)\sqrt{\frac{1}{\mathbb{E}\left[\tilde{X}(t)\right]}} - a_2\sqrt{\mathbb{E}\left[\tilde{X}(t)\right]}\right)^2 dt$$

$$+a_1\int_0^1 \varphi^\top(t)\varphi(t)\left(\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]-\frac{1}{\mathbb{E}\left[\tilde{X}(t)\right]}\right)dta_1.$$

Since from (3.3.4), $\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]-\frac{1}{\mathbb{E}[\tilde{X}(t)]}$ is strictly positive, $a^\top\Sigma a=0$ if and only if

$$\begin{cases}\int_0^1\left(a_1\varphi^\top(t)\sqrt{\frac{1}{\mathbb{E}\left[\tilde{X}(t)\right]}}-a_2\sqrt{\mathbb{E}\left[\tilde{X}(t)\right]}\right)^2dt=0,\\ \int_0^1(a_1\varphi^\top(t))^2\left(\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]-\frac{1}{\mathbb{E}\left[\tilde{X}(t)\right]}\right)dt=0.\end{cases}$$

$\int_0^1(a_1\varphi^\top(t))^2\left(\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]-\frac{1}{\mathbb{E}\left[\tilde{X}(t)\right]}\right)dt=0$ if and only if $a_1\varphi^\top(t)=0$. By Assumption 2.2, $a_1=0$. Further,

$$\int_0^1\left(a_1\varphi^\top(t)\sqrt{\frac{1}{\mathbb{E}\left[\tilde{X}(t)\right]}}-a_2\sqrt{\mathbb{E}\left[\tilde{X}(t)\right]}\right)^2dt=0 \text{ and } a_1=0$$

imply that $a_2=0$. This is $a\Sigma a^\top=0$ if and only if $a=[a_1^\top,a_2]$ is a zero vector. This completes the proof. □

Proof of the Proposition 3.3.9. **(1)** By the fact $\|\varphi(t)\|\leqslant K_\varphi$,

$$\left\|\frac{1}{T}\int_0^T\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt-\frac{1}{T}\int_0^T\frac{\varphi^\top(t)\varphi(t)}{X(t)}dt\right\|\leqslant\frac{1}{T}\int_0^T\left\|\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}-\frac{\varphi^\top(t)\varphi(t)}{X(t)}\right\|dt$$

$$\leqslant\frac{1}{T}\int_0^T\left\|\varphi^\top(t)\varphi(t)\right\|\left|\frac{1}{\tilde{X}(t)}-\frac{1}{X(t)}\right|dt\leqslant K_\varphi^2\frac{1}{T}\int_0^T\left|\frac{1}{\tilde{X}(t)}-\frac{1}{X(t)}\right|dt.$$

Further, by Proposition 3.3.7, $\frac{1}{\tilde{X}(t)}-\frac{1}{X(t)}\xrightarrow[t\to\infty]{a.s.}0$, and by the continuous version of Kronecker's Lemma, $\frac{1}{T}\int_0^T\left|\frac{1}{\tilde{X}(t)}-\frac{1}{X(t)}\right|dt\xrightarrow[t\to\infty]{a.s.}0$, which implies that

$$\frac{1}{T}\int_0^T\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt-\frac{1}{T}\int_0^T\frac{\varphi^\top(t)\varphi(t)}{X(t)}dt\xrightarrow[t\to\infty]{a.s.}0. \tag{B.33}$$

Next step is to prove the convergence of $\frac{1}{T}\int_0^T\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt$. First, since the period is supposed to be 1,

$$\frac{1}{T}\int_0^T\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt=\frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_{i-1}^{i}\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt+\frac{1}{T}\int_{\lfloor T\rfloor}^{T}\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt.$$

According to Proposition 3.3.3, the sequence of random variables $\{\tilde{X}(u+k-1)\}_{k\in\mathbb{N}_+}$ is stationary and ergodic. Since the function $y = 1/x, x > 0$ is a measurable function, *Birkhoff Ergodic Theorem* is applied to the process $\left\{\int_{i-1}^{i} \frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt\right\}_{i\in\mathbb{N}_+}$.

$$\frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_{i-1}^{i}\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \varphi^\top(u)\varphi(u)\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]du. \tag{B.34}$$

Further, using *Birkhoff Ergodic Theorem* again,

$$\frac{1}{T}\int_0^{\lfloor T\rfloor+1}\frac{1}{\tilde{X}(t)}dt = \frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor+1}\int_{i-1}^{i}\frac{1}{\tilde{X}(t)}dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]du.$$

Similarly, $\frac{1}{T}\int_0^{\lfloor T\rfloor}\frac{1}{\tilde{X}(t)}dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]du$, which implies that

$$\frac{1}{T}\int_{\lfloor T\rfloor}^{\lfloor T\rfloor+1}\frac{1}{\tilde{X}(t)}dt = \frac{1}{T}\int_0^{\lfloor T\rfloor+1}\frac{1}{\tilde{X}(t)}dt - \frac{1}{T}\int_0^{\lfloor T\rfloor}\frac{1}{\tilde{X}(t)}dt \xrightarrow[T\to\infty]{a.s.} 0. \tag{B.35}$$

From Assumption 2.2, $\left\|\frac{1}{T}\int_{\lfloor T\rfloor}^{\lfloor T\rfloor+1}\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt\right\| \leqslant K_\varphi^2\frac{1}{T}\left|\int_{\lfloor T\rfloor}^{\lfloor T\rfloor+1}\frac{1}{\tilde{X}(t)}dt\right|$, and (B.35) implies

$$\frac{1}{T}\int_{\lfloor T\rfloor}^{\lfloor T\rfloor+1}\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt \xrightarrow[T\to\infty]{a.s.} 0. \tag{B.36}$$

(B.33), (B.34) and (B.36) finish the proof of Step (1):

$$\frac{1}{T}\int_0^{T}\frac{\varphi^\top(t)\varphi(t)}{\tilde{X}(t)}dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \varphi^\top(u)\varphi(u)\mathbb{E}\left[\frac{1}{\tilde{X}(t)}\right]du. \tag{B.37}$$

Further,

$$\frac{1}{T}\int_0^T\varphi(t)dt = \frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_{i-1}^{i}\varphi(t)dt + \frac{1}{T}\int_{\lfloor T\rfloor}^{T}\varphi(t)dt.$$

By the periodic property of the function $\varphi(t)$, let $u = t-(i-1)$ with $u \in [0,1)$, in accordance with Assumption 2.2, $\varphi\left(u+(i-1)\right) = \varphi(u)$,

$$\frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_{i-1}^{i}\varphi(t)dt = \frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_0^1\varphi(u)du \xrightarrow[T\to\infty]{} \int_0^1\varphi(u)du. \tag{B.38}$$

Further, by letting $u = t-\lfloor T\rfloor$ with $u \in [0,1]$, by to Assumption 2.2, $\varphi(u+\lfloor T\rfloor) = \varphi(u)$,

$$\left\|\frac{1}{T}\int_{\lfloor T\rfloor}^{T}\varphi(t)dt\right\| \leqslant \frac{1}{T}\int_{\lfloor T\rfloor}^{\lfloor T\rfloor+1}\|\varphi(t)\|\,dt \leqslant \frac{1}{T}K_\varphi \xrightarrow[T\to\infty]{} 0. \tag{B.39}$$

(B.38) and (B.39) imply that

$$\frac{1}{T}\int_0^T \varphi(t)dt \xrightarrow[T\to\infty]{} \int_0^1 \varphi(u)du. \tag{B.40}$$

From Proposition 3.3.4, $\tilde{X}(t) - X(t) \xrightarrow[t\to\infty]{a.s.\text{ and } L^m} 0$, and by the continuous version of Kronecker's Lemma,

$$\frac{1}{T}\int_0^T (\tilde{X}(t) - X(t))dt \xrightarrow[T\to\infty]{a.s.\text{ and } L^m} 0.$$

Then,

$$\frac{1}{T}\int_0^T \tilde{X}(t)dt = \frac{1}{T}\int_0^{\lfloor T\rfloor} \tilde{X}(t)dt + \frac{1}{T}\int_{\lfloor T\rfloor}^T \tilde{X}(t)dt = \frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_{i-1}^i \tilde{X}(t)dt + \frac{1}{T}\int_{\lfloor T\rfloor}^T \tilde{X}(t)dt$$

From the fact $0 \leqslant \frac{1}{T}\int_{\lfloor T\rfloor}^T \tilde{X}(t)dt \leqslant \frac{1}{T}\int_{\lfloor T\rfloor}^{\lfloor T\rfloor+1} \tilde{X}(t)dt$,

$$\frac{1}{T}\int_0^{\lfloor T\rfloor} \tilde{X}(t)dt \leqslant \frac{1}{T}\int_0^T \tilde{X}(t)dt \leqslant \frac{1}{T}\int_0^{\lfloor T\rfloor+1} \tilde{X}(t)dt.$$

Since the process $\{\tilde{X}(t+k-1), t \in [0,1]\}_{k\in\mathbb{N}_+}$ is stationary and ergodic, by using *Birkhoff Ergodic Theorem*,

$$\frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor}\int_{i-1}^i \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \mathbb{E}\left[\tilde{X}(t)\right]dt. \tag{B.41}$$

Similarly,

$$\frac{1}{T}\sum_{i=1}^{\lfloor T\rfloor+1}\int_{i-1}^i \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \mathbb{E}\left[\tilde{X}(t)\right]dt. \tag{B.42}$$

(B.41) and (B.42) imply that

$$\frac{1}{T}\int_0^T \tilde{X}(t)dt \xrightarrow[T\to\infty]{a.s.} \int_0^1 \mathbb{E}\left[\tilde{X}(t)\right]dt \tag{B.43}$$

From (B.37), (B.40), and (B.43), one concludes that $\frac{1}{T}Q_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \Sigma$.

(2) From Proposition 3.3.8, Σ is invertible. Then, $TQ_{[0,T]}^{-1} \xrightarrow[T\to\infty]{a.s.} \Sigma^{-1}$, this completes the proof. □

Proof of Proposition 3.3.10. **(1)** For $i = 1, 2, \cdots$, by Itô isometry,

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T \frac{\varphi_i(t)}{\sqrt{X(t)}}dB_t\right)^2\right] = \mathbb{E}\left[\frac{1}{T}\int_0^T \frac{\varphi_i^2(t)}{X(t)}dt\right] = \frac{1}{T}\int_0^T \varphi_i^2(t)\mathbb{E}\left[\frac{1}{X(t)}\right]dt.$$

By Assumption 2.2 and (B.32), $\varphi_i^2(t) \leqslant K_\varphi^2$, and

$$\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[r_t^{-1}\right] < \infty,$$

which implies that

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T \frac{\varphi_i(t)}{\sqrt{X(t)}}dB_t\right)^2\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right]K_\varphi^2\frac{1}{T}\int_0^T dt = \sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right]K_\varphi^2.$$

For the last entry of $\frac{1}{\sqrt{T}}W_{[0,T]}$, to prove the boundedness of $\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T \sqrt{X(t)}dB_t\right)^2\right]$. Indeed,

$$\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T \sqrt{X(t)}dB_t\right)^2\right] = \mathbb{E}\left[\frac{1}{T}\int_0^T X(t)dt\right].$$

By Proposition 3.2.2, $\sup_{t\geqslant 0}\mathbb{E}[|X(t)|] < \infty$, then, $\mathbb{E}\left[\frac{1}{T}\int_0^T X(t)dt\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}[|X(t)|] < \infty$. This proves Part **(1)**.

(2) For any $\varepsilon > 0$,

$$\mathrm{P}\left(\sup_{2^k\leqslant T\leqslant 2^{k+1}}\left\|\frac{1}{T}W_{[0,T]}\right\| > \varepsilon\right) \leqslant \mathrm{P}\left(\sup_{2^k\leqslant T\leqslant 2^{k+1}}\left|W_{[0,T]}(i)\right| > \varepsilon 2^k\right)$$

$$\leqslant \frac{\mathbb{E}\left[\sup_{2^k\leqslant T\leqslant 2^{k+1}}\left\|W_{[0,T]}\right\|^2\right]}{\varepsilon^2 2^{2k}} \leqslant \max\left(\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right]K_\varphi^2, \sup_{t\geqslant 0}\mathbb{E}[|X(t)|]\right)\frac{2^{k+1}}{\varepsilon^2 2^{2k}},$$

and then

$$\sum_{k=1}^{\infty}\mathrm{P}\left(\sup_{2^k\leqslant T\leqslant 2^{k+1}}\left\|W_{[0,T]}\right\| > \varepsilon\right) \leqslant \max\left(\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right]K_\varphi^2, \sup_{t\geqslant 0}\mathbb{E}[|X(t)|]\right)\frac{4}{\varepsilon^2} < \infty.$$

Therefore, by Borel–Cantelli's lemma, $\frac{1}{T}W_{[0,T]} \xrightarrow[T\to\infty]{a.s.} 0$. □

Proof of Lemma 3.3.1. **(1).** The proof is proven by combining Proposition 3.3.4, Proposition 3.3.7 along with Markov's inequality and Itô's isometry.

(2). Weak convergence on the space $C_{p+1}[0, 1]$ with the uniform topology is concerned;

$C_{p+1}[0,1]$ is metrized by taking the distance between two functions $x = x(t)$ and $y = y(t)$ to be

$$\rho(x, y) = \sup_{0 \leqslant t \leqslant 1} \|x(t) - y(t)\|.$$

At each $t \in [0,1]$, $f : [0,1] \mapsto \mathbb{R}^{p+1}$. $\|f\| = \sup_{0 \leqslant s \leqslant 1} \|f(s)\|$. The modulus of continuity of an arbitrary function f on $[0,1]$ is defined as

$$w_f(\delta) = \sup_{|s-u| \leqslant \delta} \|f(s) - f(u)\|, \quad 0 < \delta \leqslant 1. \tag{B.44}$$

A necessary and sufficient condition for f to be uniformly continuous over $[0,1]$ is $\lim_{\delta \to 0} w_f(\delta) = 0$. In our case, for $0 < \delta \leqslant 1$,

$$\lim_{\delta \to 0} w_{\widetilde{W}^{(T)}}(\delta) = \lim_{\delta \to 0} \sup_{|s-u| \leqslant \delta} \left\| \widetilde{W}^{(T)}(s) - \widetilde{W}^{(T)}(u) \right\|.$$

For any $\varepsilon > 0$, without loss of generality, suppose $u < s$. By Markov inequality,

$$\mathrm{P}\left(\sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left\| \widetilde{W}^{(T)}(s) - \widetilde{W}^{(T)}(u) \right\|^2 > \varepsilon \right)$$

$$= \mathrm{P}\left(\sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left(\sum_{i=1}^{p} \left(\int_{uT}^{sT} \left(\varphi_i(t) / \sqrt{\tilde{X}(t)} \right) dB_t \right)^2 + \left(\int_{uT}^{sT} \sqrt{\tilde{X}(t)} dB_t \right)^2 \right) > \varepsilon T \right).$$

By Markov's inequality,

$$\mathrm{P}\left(\sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left\| \widetilde{W}^{(T)}(s) - \widetilde{W}^{(T)}(u) \right\|^2 > \varepsilon \right)$$

$$\leqslant \frac{\mathbb{E}\left[\sum_{i=1}^{p} \sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left(\int_{uT}^{sT} \left(\varphi_i(t) / \sqrt{\tilde{X}(t)} \right) dB_t \right)^2 \right] + \mathbb{E}\left[\sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left(\int_{uT}^{sT} \sqrt{\tilde{X}(t)} dB_t \right)^2 \right]}{\varepsilon T}$$

$$= \frac{\sum_{i=1}^{p} \mathbb{E}\left[\sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left(\int_{uT}^{sT} \left(\varphi_i(t) / \sqrt{\tilde{X}(t)} \right) dB_t \right)^2 \right] + \mathbb{E}\left[\sup_{\frac{1}{2^{k+1}} < s-u \leqslant \frac{1}{2^k}} \left(\int_{uT}^{sT} \sqrt{\tilde{X}(t)} dB_t \right)^2 \right]}{\varepsilon T}.$$

From Burkholder-Davis-Gundy inequality,

$$\mathrm{P}\left(\sup_{\frac{1}{2^{k+1}}<s-u\leqslant\frac{1}{2^k}}\left\|\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right\|^2>\varepsilon\right)\leqslant 4\left(pK_\varphi^2\sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right]+\sup_{t\geqslant 0}\mathbb{E}\left[\tilde{X}(t)\right]\right)\frac{1}{2^k\varepsilon}.$$

Then,

$$\sum_{k=1}^{\infty}\mathrm{P}\left(\sup_{\frac{1}{2^{k+1}}<s-u\leqslant\frac{1}{2^k}}\left\|\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right\|^2>\varepsilon\right)$$

$$\leqslant\left(pK_\varphi^2\sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right]+\sup_{t\geqslant 0}\mathbb{E}\left[\tilde{X}(t)\right]\right)\frac{4}{\varepsilon}\sum_{k=1}^{\infty}\frac{1}{2^k}<\infty.$$

By Borel–Cantelli Lemma, $\lim_{\delta\to 0}\sup_{|s-u|\leqslant\delta}\left\|\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right\|^2=0$, which implies that $\lim_{\delta\to 0}w_{\widetilde{W}^{(T)}}(\delta)=0$. Then, $\widetilde{W}^{(T)}(s)$ is uniformly continuous with respect to s, i.e. $\widetilde{W}^{(T)}(s)\in C_{p+1}[0,1]$. Let P_T be the probability measure of the functional diffusion process $\widetilde{W}^{(T)}(s)$ defined on the space $C_{p+1}[0,1]$. From Theorem 7.3 in Billingsley [2013], $(\mathrm{P}_T)_{T>0}$ is tight if and only if for each positive $\eta>0$, there exist an a and an T_0 such that

$$\mathrm{P}_T\left(\left\{\widetilde{W}^{(T)}\in C_{p+1}[0,1]:\left\|\widetilde{W}^{(T)}(0)\right\|\geqslant a\right\}\right)\leqslant\eta,\quad T\geqslant T_0 \tag{B.45}$$

and for $\forall\,\varepsilon>0$,

$$\lim_{\delta\to 0}\limsup_{T\to\infty}\mathrm{P}_T\left(\left\{\widetilde{W}^{(T)}\in C_{p+1}[0,1]:w_{\widetilde{W}}(\delta)\geqslant\varepsilon\right\}\right)=0. \tag{B.46}$$

The relation in (B.45) is established by combining Itô's isometry and Markov's inequality along with some algebraic computations. Further, by Markov inequality, for $\forall\,\varepsilon>0$,

$$\mathrm{P}_T\left(\left\{\widetilde{W}^{(T)}\in C_{p+1}[0,1]:w_{\widetilde{W}}(\delta)\geqslant\varepsilon\right\}\right)\leqslant\mathbb{E}\left[\sup_{|s-u|\leqslant\delta}\left\|\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right\|^2\right]\Big/\varepsilon^2.$$

Without loss of generality, suppose that $s>u$. Then,

$$\mathbb{E}\left[\sup_{|s-u|\leqslant\delta}\left\|\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right\|^2\right]\leqslant\frac{1}{T}\sum_{i=1}^{p}\mathbb{E}\left[\sup_{s-u\leqslant\delta}\left(\int_{uT}^{sT}(\varphi_i(t)/\sqrt{\tilde{X}(t)})dB_t\right)^2\right]$$

$$+\frac{1}{T}\mathbb{E}\left[\sup_{s-u\leqslant\delta}\left(\int_{uT}^{sT}\sqrt{\tilde{X}(t)}dB_t\right)^2\right].$$

By Burkholder-Davis-Gundy inequality,

$$\mathbb{E}\left[\sup_{s-u\leqslant\delta}\left(\int_{uT}^{sT}(\varphi_i(t)/\sqrt{\tilde{X}(t)})dB_t\right)^2\right]\leqslant 4K_\varphi^2\int_{uT}^{(u+\delta)T}\mathbb{E}\left[1/\tilde{X}(t)\right]dt\leqslant 4K_\varphi^2\sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right]\delta T.$$

Similarly, $\mathbb{E}\left[\sup_{s-u\leqslant\delta}\left(\int_{uT}^{sT}\sqrt{\tilde{X}(t)}dB_t\right)^2\right]\leqslant 4\sup_{t\geqslant 0}\mathbb{E}\left[\tilde{X}(t)\right]\delta T$. Therefore,

$$\mathbb{E}\left[\sup_{|s-u|\leqslant\delta}\left\|\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right\|^2\right]\leqslant 4\left(\sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right]+\sup_{t\geqslant 0}\mathbb{E}\left[\tilde{X}(t)\right]\right)\delta.$$

The fact $\sup_{t\geqslant 0}\mathbb{E}\left[1/\tilde{X}(t)\right]<\infty$ and $\sup_{t\geqslant 0}\mathbb{E}\left[\tilde{X}(t)\right]<\infty$, for $\forall\,\varepsilon>0$, implies

$$\lim_{\delta\to 0}\limsup_{T\to\infty}\mathrm{P}_T\left(\left\{\widetilde{W}^{(T)}\in C_{p+1}[0,1]:w_{\widetilde{W}}(\delta)\geqslant\varepsilon\right\}\right)=0.$$

This completes the proof of Part **(2)**.

(3) $\widetilde{W}^{(T)}(s)=\frac{1}{\sqrt{T}}\sum_{i=1}^{\lfloor sT\rfloor}\int_{i-1}^{i}\left((\varphi(t),-\tilde{X}(t))/\sqrt{\tilde{X}(t)}\right)dB_t+\frac{1}{\sqrt{T}}\int_{\lfloor sT\rfloor}^{sT}\left((\varphi(t),-\tilde{X}(t))/\sqrt{\tilde{X}(t)}\right)dB_t$, where $\frac{1}{\sqrt{T}}\int_{\lfloor sT\rfloor}^{sT}\left((\varphi(t),-\tilde{X}(t))/\sqrt{\tilde{X}(t)}\right)dB_t\xrightarrow[T\to\infty]{\mathrm{P}}0$. It is obvious that the random sequence $\{\vartheta_i=\int_{i-1}^{i}\frac{(\varphi(t),-\tilde{X}(t))}{\sqrt{\tilde{X}(t)}}dB_t,i\geqslant 1\}$ is a strictly stationary and ergodic sequence. Further, $\mathbb{E}[\vartheta_i|\mathscr{F}_{i-1}]=0$ and $\mathrm{Var}(\vartheta_i)=\Sigma$, and $\|\Sigma\|<\infty$. Then, by combining central limit theorem for martingale difference sequences and Slutsky's theorem, $\widetilde{W}^{(T)}(s)\xrightarrow[T\to\infty]{\mathrm{D}}\widetilde{W}^*(s)\sim\mathcal{N}_{p+1}(0,s\Sigma)$, Further, for $0<u<s<1$, one has

$$\begin{aligned}\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)&=\frac{1}{\sqrt{T}}\sum_{i=\lfloor uT\rfloor+1}^{\lfloor sT\rfloor}\int_{i-1}^{i}\frac{(\varphi(t),-\tilde{X}(t))}{\sqrt{\tilde{X}(t)}}dB_t-\frac{1}{\sqrt{T}}\int_{\lfloor uT\rfloor}^{uT}\frac{(\varphi(t),-\tilde{X}(t))}{\sqrt{\tilde{X}(t)}}dB_t\\&\quad+\frac{1}{\sqrt{T}}\int_{\lfloor sT\rfloor}^{sT}\frac{(\varphi(t),-\tilde{X}(t))}{\sqrt{\tilde{X}(t)}}dB_t.\end{aligned}$$

One proves that

$$\frac{1}{\sqrt{T}}\int_{\lfloor uT\rfloor}^{uT}\frac{(\varphi(t),-\tilde{X}(t))}{\sqrt{\tilde{X}(t)}}dB_t\xrightarrow[T\to\infty]{\mathrm{P}}0\ \text{ and }\ \frac{1}{\sqrt{T}}\int_{\lfloor sT\rfloor}^{sT}\frac{(\varphi(t),-\tilde{X}(t))}{\sqrt{\tilde{X}(t)}}dB_t\xrightarrow[T\to\infty]{\mathrm{P}}0,$$

$$\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\xrightarrow[T\to\infty]{\mathrm{D}}\tilde{W}^*(s-u)\sim\mathcal{N}_{p+1}(0,(s-u)\Sigma),$$

$$\mathrm{Cov}\left(\widetilde{W}^{(T)}(u),\left(\widetilde{W}^{(T)}(s)-\widetilde{W}^{(T)}(u)\right)\right)=\mathbf{0}.$$

This results and Part **(1)** of the lemma imply that $\{\widetilde{W}^{(T)}(s),0\leqslant s\leqslant 1\}$ converges weakly to a mean zero Gaussian process $\{\widetilde{W}^*(s), 0 \leqslant s \leqslant 1\}$ with $\mathrm{Cov}\left(\widetilde{W}^*(s), \widetilde{W}^*(u)\right) = \min(u, s)\Sigma$ for all $0 \leqslant u \neq s \leqslant 1$. This completes the proof. □

Proof of Proposition 3.3.11. First, Proposition 3.3.9 and Proposition 3.3.10 imply the following convergence: $\hat{\theta}_T \xrightarrow[T\to\infty]{a.s.} \theta$. Further, $\rho_T = \sigma\sqrt{T}Q_{[0,T]}^{-1}W_{[0,T]} = \sigma T Q_{[0,T]}^{-1}\frac{1}{\sqrt{T}}W_{[0,T]}$. By Proposition 3.3.9, $\sigma T Q_{[0,T]}^{-1} \xrightarrow[T\to\infty]{a.s.} \sigma\Sigma^{-1}$, and by Lemma 3.3.1,

$$\frac{1}{\sqrt{T}}W_{[0,T]} \xrightarrow[T\to\infty]{\mathrm{D}} \widetilde{W}_1^* \sim \mathcal{N}_{p+1}(0, \Sigma). \tag{B.47}$$

Then, by Slutsky's Theorem, $\rho_T = \sigma T Q_{[0,T]}^{-1}\frac{1}{\sqrt{T}}W_{[0,T]} \xrightarrow[T\to\infty]{\mathrm{D}} \sigma\Sigma^{-1}W_1^* = \rho$. Note that Σ^{-1} is non-random and symmetric, then, $\rho \sim \mathcal{N}_{p+1}\left(0, \sigma^2\Sigma^{-1}\right)$. Hence, $\sqrt{T}(\hat{\theta}_T - \theta) \xrightarrow[T\to\infty]{\mathrm{D}} \rho \sim \mathcal{N}_{p+1}\left(0, \sigma^2\Sigma^{-1}\right)$. This completes the proof. □

Proof of Theorem 3.3.1. For every $\theta \in \Theta$, and arbitrary bounded sequences $\boldsymbol{h} \in \mathbb{R}^{p+1}$,

$$\log(Z_T(h)) = \frac{1}{\sigma^2}\int_0^T \frac{S(t, \theta + \boldsymbol{h}/\sqrt{T}, X(t)) - S(t, \theta, X(t))}{X(t)}dX(t)$$

$$-\frac{1}{2\sigma^2}\int_0^T \frac{\left(S(t, \theta + \boldsymbol{h}/\sqrt{T}, X(t)) - S(t, \theta, X(t))\right)^2}{X(t)}dt,$$

then, by using the fact that $B_t = \int_0^t \frac{1}{\sigma\sqrt{X(s)}}dX(s)$ is a $\mathscr{F}_t$ measurable Brownian motion,

$$\log(Z_T(h)) = \frac{1}{\sigma}\int_0^T \frac{S(t, \theta + \boldsymbol{h}/\sqrt{T}, X(t)) - S(t, \theta, X(t))}{\sqrt{X(t)}}dB_t$$

$$-\frac{1}{2\sigma^2}\int_0^T \frac{\left(S(t, \theta + \boldsymbol{h}/\sqrt{T}, X(t)) - S(t, \theta, X(t))\right)^2}{X(t)}dt.$$

Then,

$$\log(Z_T(\boldsymbol{h})) = \frac{1}{\sigma}\int_0^T \frac{(\theta + \boldsymbol{h}/\sqrt{T})^\top(\varphi(t), -X(t))^\top - \theta^\top(\varphi(t), -X(t))^\top}{\sqrt{X(t)}}dB_t$$

$$-\frac{1}{2\sigma^2}\int_0^T \frac{\left((\theta + \boldsymbol{h}/\sqrt{T})^\top(\varphi(t), -X(t))^\top - \theta^\top(\varphi(t), -X(t))^\top\right)^2}{X(t)}dt.$$

This gives

$$\log(Z_T(\boldsymbol{h})) = \frac{1}{\sigma}\boldsymbol{h}^\top \frac{1}{\sqrt{T}} \int_0^T \frac{(\varphi(t), -X(t))^\top}{\sqrt{X(t)}} dB_t$$

$$-\frac{1}{2\sigma^2}\boldsymbol{h}^\top \left(\frac{1}{T}\int_0^T \frac{(\varphi(t), -X(t))^\top(\varphi(t), -X(t))}{X(t)} dt\right)\boldsymbol{h}.$$

Let $\Delta_T(\theta_0, X^T) = \frac{1}{\sqrt{T}} W_{[0,T]}$ and $r_T(\theta_0, \boldsymbol{h}, X^T) = \frac{1}{2\sigma^2}\boldsymbol{h}^\top \left(\frac{1}{T} Q_{[0,T]} - \Sigma\right)\boldsymbol{h}$. Then, together with (3.3.1) and (3.3.2), $\log(Z_T(\boldsymbol{h})) = \frac{1}{\sigma}\boldsymbol{h}^\top \Delta_T(\theta_0, X^T) - \frac{1}{2\sigma^2}\boldsymbol{h}^\top\Sigma\boldsymbol{h} - r_T(\theta_0, \boldsymbol{h}, X^T)$, where $\Delta_T(\theta_0, X^T) = \frac{1}{\sqrt{T}} W_{[0,T]} \xrightarrow[T\to\infty]{\mathrm{D}} \widetilde{W}^*(1) \sim \mathcal{N}_{p+1}(0, \Sigma)$, and by Proposition 3.3.9 $\frac{1}{T} Q_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \Sigma$, which implies that $r_T(\theta_0, \boldsymbol{h}, X^T) \xrightarrow[T\to\infty]{\mathrm{P}_{\theta_0}} 0$. This completes the proof. □

B.4 Proofs on asymptotic test and distributional risk analysis

B.4.1 On convergence of the test statistic

Proof of Proposition 3.4.1. From Proposition 3.3.12 and Proposition 3.3.9 along with Slutsky's theorem, $\varsigma_T \xrightarrow[T\to\infty]{D} \varsigma \sim \mathcal{N}_{p+1}(G^* r_0, \sigma^2 G^* M \Sigma^{-1})$. Furthermore, by Proposition 3.3.9,

$\psi_T = \varsigma_T^\top \hat{\Gamma} \varsigma_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi = \varsigma^\top \Gamma \varsigma$, where $\varsigma_T \xrightarrow[T\to\infty]{D} \varsigma \sim \mathcal{N}_{p+1}(G^* r_0, \sigma^2 G^* M \Sigma^{-1})$. Further, the proof follows from Theorem 5.1.3 in Mathai and Provost [1992], which is similar to the proof of Proposition 2.4.9. This completes the proof. □

C Proofs related to GCKLS process

C.1 Boundary classification for regular diffusion processes

From Karlin and Taylor [1981], let $X(t)$ be a regular diffusion process on an interval $I = (0, +\infty)$ where 0 is the left boundary and ∞ is the right boundary. From [Karlin and Taylor, 1981, Chapter 15, Section 1, Page 159], let $\Delta_h X(t)$ be the increment in the process accrued over a time interval of length h. Thus, $\Delta_h X(t) = X_{t+h} - X(t)$. let

$$\lim_{h\to 0^+} \frac{1}{h}\mathbb{E}\left[\Delta_h X(t)|X(t) = x\right] = \mu(x,t) = L(t) - \alpha x,$$

and

$$\lim_{h\to 0^+} \frac{1}{h}\mathbb{E}\left[\{\Delta_h X(t)\}^2|X(t) = x\right] = \sigma^2(x,t) = \sigma^2(x)^{2\delta}.$$

For x in $(0,\infty)$, we postulate the continuous infinitesimal drift and variance coefficients $L(t) - \alpha x$ and $\sigma^2(x)^{2\delta} > 0$, respectively. In the following, we concentrate on the left boundary 0, the right being entirely similar. Let T_z is the hitting time to z and $T_{a,b} = T_a \wedge T_b = \min\{T_a, T_b\}$. The approach is to let a decrease to 0 in the quantities

$$u(x) = u_{a,b}(x) = \Pr(T_b < T_a|X(0) = x) \tag{C.1}$$

and

$$v(x) = v_{a,b}(x) = \mathbb{E}[T_{a,b}|X(0) = x], \quad 0 < a < x < b < \infty. \tag{C.2}$$

The scale function $\boldsymbol{Q}(x,t)$ is defined with explicit expression

$$\boldsymbol{Q}(x,t) = \int_{x_0}^{x} \boldsymbol{q}(u,t)du, \quad \boldsymbol{q}(u,t) = \exp\left\{-\int_{u_0}^{u} 2(L(t) - \alpha\eta)/(\sigma^2\eta^{2\delta})d\eta\right\} \tag{C.3}$$

where x_0 and u_0 are arbitrary fixed points in the open interval $(0,\infty)$. It simplifies the exposition if we introduce the scale measure, the function $\boldsymbol{Q}([J],t)$ of closed intervals $J = [c,d] \subset (0,\infty)$ defined by

$$\boldsymbol{Q}([J],t) = \boldsymbol{Q}([c,d],t) = \boldsymbol{Q}(d,t) - \boldsymbol{Q}(c,t).$$

let both the scale function and scale measure by the same symbol $\boldsymbol{Q}$; no confusion results.

We freely use the scale measure $\frac{\partial}{\partial x}\boldsymbol{Q}(x,t)dx = \boldsymbol{Q}([dx],t)$ of an infinitesimal interval $[x, x+dx]$ with $\boldsymbol{Q}([dx],t) = \boldsymbol{Q}(x+dx,t) - \boldsymbol{Q}(x,t) = \frac{\partial}{\partial x}\boldsymbol{Q}(x,t)dx = q(x,t)dx$. For example, we evaluate $\int_c^d f(x)\frac{\partial}{\partial x}\boldsymbol{Q}(x,t)dx$ using the usual integral $\int_c^d f(x)q(x,t)dx$, for, say, piecewise continuous function $f(x)$. It is easy to check that $0 < \boldsymbol{Q}([c,d],t) < \infty$ for

$0 < c < d < \infty$, and that

$$\boldsymbol{Q}([c,d],t) = \boldsymbol{Q}([c,x],t) + \boldsymbol{Q}([x,d],t), \quad \text{for} \quad 0 < c < x < d < \infty. \tag{C.4}$$

Similarly, we introduce the speed measure $\mathcal{M}$ induced by the speed density $\boldsymbol{m}(x,t) = 1/(\sigma^2 x^{2\delta}\boldsymbol{q}(x,t))$, where

$$\mathcal{M}([J],t) = \mathcal{M}([c,d],t) = \int_c^d \boldsymbol{m}(x,t)dx, \quad J = [c,d] \subset (0,\infty).$$

Again, $\mathcal{M}([J],t)$ is positive and finite for $J = [a,b] \subset (0,\infty)$. In terms of the scale and speed measures, (C.1) and (C.2) are written

$$u(x,t) = u_{a,b}(x,t) = \boldsymbol{Q}([a,x],t)/\boldsymbol{Q}([a,b],t), \quad 0 < a < x < b < \infty. \tag{C.5}$$

and

$$\begin{aligned} v(x,t) &= v_{a,b}(x,t) = 2\Big\{u(x,t)\textstyle\int_x^b \boldsymbol{Q}([\eta,b],t)\frac{\partial}{\partial\eta}\mathcal{M}(\eta,t)d\eta \\ &\quad +[1-u(x,t)]\textstyle\int_a^x \boldsymbol{Q}([a,\eta],t)\frac{\partial}{\partial\eta}\mathcal{M}(\eta,t)d\eta\Big\}. \end{aligned} \tag{C.6}$$

It follows from the nonnegativity of the measure $\boldsymbol{Q}$ and (C.4) that $\boldsymbol{Q}([a,b],t)$ is monotonic in a for fixed b and that therefore we may define $\boldsymbol{Q}((0,b],t) \leqslant \infty$ by

$$\boldsymbol{Q}((0,b],T) = \lim_{a\downarrow 0} \boldsymbol{Q}([a,b],t) \leqslant \infty, \quad 0 < b < \infty. \tag{C.7}$$

If $[a,b] \subset (0,\infty)$, then, $0 \leqslant \boldsymbol{Q}([a,b],t) < \infty$. As an easy consequence of this and (C.4), $\boldsymbol{Q}((0,b],t) = \infty$ for some $b \in (0,\infty)$ if and only if

$$\boldsymbol{Q}((0,b],t) = \infty \quad \text{for all } b \in (0,\infty) \tag{C.8}$$

Because the limiting behavior occurs as $a \to 0^+$, we don't consider the limit of right endpoint. Then, if $\boldsymbol{Q}((0,b],t) = \infty$, $\boldsymbol{Q}((0,b],t) = \infty \quad$ for all $b \in (0,\infty)$.

Definition C.1 (Definition 6.1 [Karlin and Taylor, 1981] Page 228)**.** The boundary 0 is *attracting* if $\boldsymbol{Q}((0,x],t) < \infty$ and this criterion applies independently of x in $(0,\infty)$.

We define

$$\boldsymbol{\Sigma}(0,t) = \lim_{a\downarrow 0}\int_a^x \boldsymbol{Q}([a,\xi],t)\frac{\partial}{\partial\xi}\mathcal{M}(\xi,t)d\xi = \int_0^x \boldsymbol{Q}([a,\xi],t)\frac{\partial}{\partial\xi}\mathcal{M}(\xi,t)d\xi$$

$$= \int_0^x\left\{\int_0^\xi q(\eta,t)d\eta\right\}\boldsymbol{m}(\xi,t)d\xi = \int_0^x\left\{\int_0^\xi \boldsymbol{m}(\xi,t)d\xi\right\}q(\eta,t)d\eta \tag{C.9}$$

Note that introduced the notation $\boldsymbol{\Sigma}(0,t) = \boldsymbol{\Sigma}(0)$ to represent the above double integral. It depends on t but in later considerations only whether its value is finite or infinite is relevant and we can therefore suppress the dependence on t without ambiguity. Expressed in terms of $\boldsymbol{\Sigma}(0)$ the following dichotomy.

Definition C.2 (Definition 6.2 [Karlin and Taylor, 1981] Page 230)**.** The boundary 0 is said to be

1. attainable if $\boldsymbol{\Sigma}(0) < \infty$,

2. unattainable if $\boldsymbol{\Sigma}(0) = \infty$.

C.2 On the solution of SDE and derivation of ergodicity

Proposition C.1. *Proposition 2.13 (Yamada & Watanabe 1971) [Karatzas and Shreve, 1998, Page 291]. Let us suppose that the coefficients of the one-dimensional equation (d=r=1),*

$$dX(t) = \varphi(t,X(t))dt + \sigma(t,X(t))dW_t \tag{C.10}$$

satisfy the condition $|\varphi(t,x) - \varphi(t,y)| \leqslant K|x-y|,\quad$ *and* $\quad|\sigma(t,x)-\sigma(t,y)| \leqslant h(|x-y|)$, *for every* $0 \leqslant t < \infty$ *and* $x \in \mathbb{R}, y \in \mathbb{R}$, *where* K *is a positive constant and* $h : [0,\infty) \to [0,\infty)$ *is a strictly increasing function with* $h(0) = 0$ *and* $\int_{(0,\epsilon)} h^{-2}(u)du = \infty;$ for any $\epsilon > 0$. *Then strong and uniqueness hold for the solution of the SDE* (C.10).

The following content is from Revuz and Yor [1999] [Chapter III Page 79].

We consider a Markov process $\{X(t), t \geqslant 0\}$ with state space $(E, \mathscr{E})$. For $A \in \mathscr{E}$, $s < t$, the conditional probability $P[X(t) \in A|\sigma(X(u), u \leqslant s)]$ should be a function of $X(s)$, that is of the form $g(X(s))$, where g is an $\mathscr{E}-$ measurable and taking its values on the interval $[0, 1]$. It would better be written $g_{s,t}$ to indicate the dependence on s and t. On the other hand, this conditional expectation depends on A and clearly, is a function of A. We thus come to the idea that the above conditional expectation may be written $g_{s,t}(X(s), A)$, where for each A, $x \mapsto g_{s,t}(x, A)$ is measurable and for each x, $A \mapsto g_{s,t}(x, A)$ is a probability measure.

Definition C.3 (Kernel)**.** Suppose that $(E, \mathscr{E})$ and $(E_1, \mathscr{E}_1)$ are measurable spaces, a Kernel from $(E, \mathscr{E})$ to $(E_1, \mathscr{E}_1)$ is a function $\mathbb{P} : E \times \mathscr{E}_1$ such that

1. $x \mapsto \mathbb{P}(x, A)$ is a measurable function from E into $[0, +\infty)$ for each $A \in \mathscr{E}_1$.
2. $A \mapsto \mathbb{P}(x, A)$ is a positive measure on $\mathscr{E}_1$ for each $a \in E$.

If $(E, \mathscr{E}) = (E_1, \mathscr{E}_1)$, then P is said to be a kernel on $(E, \mathscr{E})$.

A kernel $\mathbb{P}$ is called a transition probability if $\mathbb{P}(x, E) = 1$ for every $x \in E$. For a positive function $f : E \mapsto \mathbb{R}_+$, we define a function $\mathbb{P}f$ on E by

$$\mathbb{P}f(x) = \int_E \mathbb{P}(x, dy) f(y).$$

Suppose that a process $\{X(t), t \geqslant 0\}$ for which, for any $s < t$, there is a transition probability $\mathbb{P}_{s,t}$ such that

$$P[X(t) \in A|\sigma(X(u), u \leqslant s)] = \mathbb{P}_{s,t}(X(s), A), \quad a.s.$$

Then for the function f defined above, $\mathbb{E}[f(X(t))|\sigma(X(u), u \leqslant s)] = \mathbb{P}_{s,t} f(X(s))$. Particularly, if the process starts at time 0 and given $X_0 = x$, let $\mathbb{P}_{0,t}(x, A)$ as $\mathbb{P}_t(x, A)$. Transition probability $\mathbb{P}_{s,t}$ is written as $\mathbb{P}_{0,t}$.

In order to prove the ergodicity of the stochastic process $\{X(t), t \geqslant 0\}$, in this paper we use the method from Höpfner et al. [2016]. For convenience, some of the preconditions needed for this proof are given below. In Proposition 4.3.1 of this paper, which proves the ergodicity, we verify that these preconditions hold one by one.

Assumption C.1. Höpfner et al. [2016]

a) For some strictly increasing sequence $(G_m)_m$ of bounded convex open sets in $\mathbb{R}^*$, and compacts $C_m := cl(G_m)$, $\mathbb{R}^* = \bigcup_{m=1}^{\infty} C_m$ and $\mathscr{E} = \mathscr{B}(\mathbb{R}^*)$.

b) $\partial(\mathbb{R}^*) \bigcap \mathbb{R}^*$ is an entrance boundary for the process $X(t)$.

b') From positions $x \in C_m \backslash G_{m+1}$ almost surely, the process $X(t)$ immediately enters G_{m+1}.

c) Defining stopping times $\mathbb{T}_m := \inf\{t > 0 : X(t) \notin C_m\}$ for the process, $\mathbb{T}_m \uparrow \infty$ as $m \to \infty$ almost surely, for every choice of a starting point $x \in \mathbb{R}^*$.

d) The components of coefficients for equation (4.2.2)

$$(t, x) \mapsto S(t, x, \theta), \quad x \mapsto \sigma(x)$$

are C^{∞}-functions on $\mathbb{R}_+ \times U$ for some open set $U \subset \mathbb{R}$ which contains $\mathbb{R}^*$.

The above assumption combines properties of the process (such as non-explosion or behavior at the boundary) with topological properties of E. In our case, $E = \mathbb{R}^*_+$.

Assumption C.2. Höpfner et al. [2016]

a) We take the drift 1-periodic in the time variable:

$$S(t, x) = S(i_1(t), x), \quad i_T(t) := t \; modulo \; 1.$$

b) a Lyapunov function $V : \mathbb{R}^* \mapsto [1, +\infty)$, in the following sense: V is $\mathscr{E}$-measurable; there is a compact K contained in $\mathbb{R}^*$ (i.e. $K \subset \mathbb{R}^*$, and $K \neq \mathbb{R}^*$) and some $\varepsilon > 0$ such that

$$P_{0,T}V \quad \textit{is bounded on } K, \;\; P_{0,T}V \leqslant V - \varepsilon \quad \textit{on } \mathbb{R}^* \backslash K.$$

By Assumption C.2 a), the semigroup of the process (4.2.2) is 1-periodic in time which means that

$$P_{s,t}(x, dy) = P_{s+k,t+k}(x, dy), \quad k \in \mathbb{N}_0.$$

This implies that the 1-skeleton chain $(X_k)_{k \in \mathbb{N}_0}$ is a time-homogeneous Markov chain. By Assumption C.2 b), it evolves as a nonnegative supermartingale as long as it stays outside K. As a consequence, the skeleton chain has to visit the compact K infinitely often, almost surely, for arbitrary choice of a starting point in $\mathbb{R}^*_+$.

The following definitions are also needed in the proof of Proposition 4.3.1 and Theorem 2.2 in Höpfner et al. [2016](page 531).

Definition C.4. [Höpfner et al., 2016, page 14] In the sequel given an SDE $dX(t) = S(t, X(t))dt + \mathbf{\Sigma}(X(t))dBt$ driven by a Brownian motion B in the Ito sense, we will need to pass to its Statonovitch form $dX(t) = \tilde{S}(t, X(t))dt + \sigma(X(t))dBt$ with **Stratonovich drift**

$$\tilde{S}(t, x) = S(t, x) - \frac{1}{2}\sigma(x)\frac{d}{dx}\sigma(x).$$

Definition C.5 (Definition 1 Höpfner et al. [2016])**.** A point x^* in $\mathbb{R}^*$ is called attainable in a sense of deterministic control if it belongs to $int(\mathbb{R}^*)$ and if the following holds: for arbitrary $x \in \mathbb{R}^*$, we can find some $\dot{h}$ in L^2_{loc} (the class of 1-dimensional measurable functions with components $\dot{h}$ satisfying $\int_0^t [\dot{h}(s)]^2 ds < \infty$ for all $t < \infty$) depending on x and x^* which drives the deterministic control system with Stratonovich drift $\tilde{S}(\cdot, \cdot)$

$$d\phi(t) = \tilde{S}(t, \phi(t))dt + \sigma(\phi(t))\dot{h}(t)dt,$$

from starting point $\varphi(0) = x$ towards the limit $x^* = \lim_{t\to\infty} \phi(t)$ under the constraint $\phi(t) \in int(E)$ for all $t > 0$. In this case we set $\phi := \phi^{(h,x,x^*)}$.

Remark C.6. The Stratonovich drift $\tilde{S}(\cdot,\cdot)$ is defined in Definition C.4.

Definition C.7 (Definition 3.3 in Höpfner et al. [2016])**.** (Page 538) We say that a point $x^* \in U \supset \mathbb{R}^*$ is of full weak Höermander dimension if there is some $N \in \mathbb{N}_+$, such that

$$\left(\dim\ \Delta_{\mathcal{L}_N^*}\right)(s, x^*) = 1 \quad independently\ of\ s \in \mathbb{R}_+$$

Assumption C.3. Höpfner et al. [2016] There is a point $x^* \in int(\mathbb{R}^*)$ with the following two properties: x^* is of full weak Höermander dimension (Definition C.7), and x^* is attainable in a sense of deterministic control.

Definition C.8. [Lie bracket of vector fields] In the mathematical field of differential topology, the Lie bracket of vector fields, also known as the Jacobi–Lie bracket or the commutator of vector fields, is an operator that assigns to any two vector fields X and Y on a smooth manifold M a third vector field denoted $[X, Y]$. If M is (an open subset of) $\mathbb{R}^n$, then the vector fields X and Y can be written as smooth maps of the form $X : M \mapsto \mathbb{R}^n$ and $Y : M \mapsto \mathbb{R}^n$, and the Lie bracket $[x, y] : M \mapsto \mathbb{R}^n$ is given by:

$$[X, Y] := \mathbb{J}_Y X - \mathbb{J}_X Y$$

where $\mathbb{J}_Y$ and $\mathbb{J}_X$ are $n \times n$ Jacobian matrices.

Definition C.9 (Definition 3.2 in Höpfner et al. [2016])**.** Define a set $\mathcal{L}$ of vector fields by **initial condition** $V_1, V_2, \cdots, V_m \in \mathcal{L}$, and arbitrary number of iteration steps

$$L \in \mathcal{L} \implies [L, V_1], [L, V_2], \cdots, [L, V_m] \in \mathcal{L} \tag{C.11}$$

For $N \in \mathbb{N}$, define the subset $\mathcal{L}_N$ by the same initial condition and at most N iterations (C.11). Write $\mathcal{L}_N^*$ for the closure of $\mathcal{L}_N$ under Lie brackets (Definition C.8); finally, write $\Delta_{\mathcal{L}_N^*} = LA(\mathcal{L}_N)$ for the linear hull of $\mathcal{L}_N^*$, i.e. the Lie algebra spanned by $\mathcal{L}_N$.

Proof of Proposition 4.2.1. **(1).** Like in the reference of [Karlin and Taylor, 1981, Chapter 15, Section 6, Page 226], define the scale function $\boldsymbol{Q}(x,t)$ with explicit expression

$$\boldsymbol{Q}(x,t) = \int_{x_0}^{x} \boldsymbol{q}(u)du, \quad \boldsymbol{q}(u,t) = \exp\left\{-\int_{u_0}^{u} 2(L(t)-\alpha\eta)/(\sigma^2\eta^{2\delta})d\eta\right\} \tag{C.12}$$

where x_0 and u_0 are arbitrary fixed points inside $(0,\infty)$. From Assumption 2.2, the function $L(t)$ is bounded in t. Then, take the scale function as an univariate function of x. Further, introduce the scale measure $d\boldsymbol{Q}(x,t) = \boldsymbol{Q}([dx],t)$ of an infinitesimal interval $[x, x+dx]$ with

$$\boldsymbol{Q}([dx],t) = \boldsymbol{Q}(x+dx,t) - \boldsymbol{Q}(x,t) = d\boldsymbol{Q}(x,t)dx = \boldsymbol{q}(x,t)dx.$$

Similarly, we introduce the speed measure $\mathcal{M}$ induced by the speed density $\boldsymbol{m}(x,t) = 1/(\sigma^2 x^{2\delta}\boldsymbol{q}(x,t))$, where

$$\mathcal{M}([c,d],t) = \int_c^d \boldsymbol{m}(x,t)dx, \quad [c,d] \subset (0,\infty).$$

To give new notations, let

$$p(l,t,x) = \int_l^x \left\{\int_\eta^x \boldsymbol{m}(\xi,t)d\xi\right\} \boldsymbol{q}(\eta,t)d\eta,$$

where x is the initial state of the process defined in (4.2.2). In our case, take $u_0 = x_0$,

$$\boldsymbol{q}(x,t) = \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}(x^{1-2\delta} - x_0^{1-2\delta}) + \frac{\alpha}{\sigma^2(1-\delta)}(x^{2-2\delta} - x_0^{2-2\delta})\right\}, \tag{C.13}$$

and

$$\boldsymbol{m}(x,t) = \frac{1}{\sigma^2 x^{2\delta}} \exp\left\{\frac{2L(t)}{\sigma^2(1-2\delta)}(x^{1-2\delta} - x_0^{1-2\delta}) - \frac{\alpha}{\sigma^2(1-\delta)}(x^{2-2\delta} - x_0^{2-2\delta})\right\}.$$

Since $0<\delta<1/2$, $1-2\delta>0$ and $2-2\delta>1$. From Assumption 2.2,
$C_{\delta,x} = \sup_{\substack{0\leqslant u\leqslant x\\ t\geqslant 0}} \exp\left\{\frac{2L(t)}{\sigma^2(1-2\delta)}(u^{1-2\delta} - x_0^{1-2\delta}) - \frac{\alpha}{\sigma^2(1-\delta)}(u^{2-2\delta} - x_0^{2-2\delta})\right\} < \infty$. Then,
$p(l,t,x) = \int_l^x \left\{\int_\eta^x \boldsymbol{m}(\xi,t)d\xi\right\}\boldsymbol{q}(\eta,t)d\eta \leqslant \int_l^x \left\{\int_\eta^x \frac{1}{\sigma^2\xi^{2\delta}} C_{\delta,x} d\xi\right\}\boldsymbol{q}(\eta,t)d\eta$, and then

$$p(l,t,x) \leqslant \frac{1}{\sigma^2} C_{\delta,x} \frac{1}{1-2\delta} x^{1-2\delta} \int_l^x \boldsymbol{q}(\eta,t)d\eta - \frac{1}{\sigma^2} C_{\delta,x} \frac{1}{1-2\delta} l^{1-2\delta} \int_l^x \boldsymbol{q}(\eta,t)d\eta.$$

This gives $p(l,t,x) \leqslant \frac{1}{\sigma^2} C_{\delta,x} \frac{1}{1-2\delta}\left(x^{1-2\delta} - l^{1-2\delta}\right) \int_l^x \boldsymbol{q}(\eta,t)d\eta$. Further,

$$C^*_{\delta,x} = \sup_{\substack{0\leqslant u\leqslant x\\ t\geqslant 0}} \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}(u^{1-2\delta} - x_0^{1-2\delta}) + \frac{\alpha}{\sigma^2(1-\delta)}(u^{2-2\delta} - x_0^{2-2\delta})\right\} < \infty,$$

which implies that $\int_l^x \boldsymbol{q}(\eta,t)d\eta \leqslant C^*_{\delta,x}(x-l)$. Then, since $\eta < x$,

$$p(l,t,x) \leqslant \frac{1}{\sigma^2} C_{\delta,x} \frac{1}{1-2\delta}\left(x^{1-2\delta} - l^{1-2\delta}\right) C^*_{\delta,x}(x-l).$$

To introduce a new notation, let $p(l,t,x) = p(l)$. Only consider the boundary when $l \to 0$ is considered. This implies that $p(0) < \infty$ for all $t \geqslant 0$ and some x in $(0,\infty)$. From Definition 6.2 in Karlin and Taylor [1981], 0 is always an attainable boundary for $0 < \delta < 1/2$.

(2). For $\delta \in (1/2, 1)$,

$$0 < 1-\delta < 1/2, \quad -1 < 1-2\delta < 0. \tag{C.14}$$

Then, (C.13) implies that

$$\boldsymbol{Q}([l,x],t) = \int_l^x \boldsymbol{q}(\eta,t)d\eta = \int_l^x \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}(\eta^{1-2\delta} - x_0^{1-2\delta}) + \frac{\alpha}{\sigma^2(1-\delta)}(\eta^{2-2\delta} - x_0^{2-2\delta})\right\} d\eta$$

$$\geqslant \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}(-x_0^{1-2\delta}) + \frac{\alpha}{\sigma^2(1-\delta)}(l^{2-2\delta} - x_0^{2-2\delta})\right\} \int_l^x \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}\eta^{1-2\delta}\right\} d\eta.$$

let $K_t^* = -\frac{2L(t)}{\sigma^2(1-2\delta)} > 0$, which is independent with x and l. Then, for η in the interval $[l,x]$,

$$\exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}\eta^{1-2\delta}\right\} = \exp\left\{K_t^*\eta^{1-2\delta}\right\} = \sum_{i=0}^{\infty} \frac{\left(K_t^*\eta^{1-2\delta}\right)^i}{i!} = \sum_{i=0}^{\infty} \frac{(K_t^*)^i}{i!}\eta^{(1-2\delta)i}.$$

Since each term $\left(K_t^*\eta^{1-2\delta}\right)^i$ is continuous for $\eta \in [l,x]$

$$\int_l^x \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}\eta^{1-2\delta}\right\} d\eta = \sum_{i=0}^{\infty} \frac{(K_t^*)^i}{i!} \int_l^x \eta^{(1-2\delta)i} d\eta$$

$$= \sum_{i=0}^{\infty} \frac{(K_t^*)^i}{i!} \frac{1}{(1-2\delta)i+1} x^{(1-2\delta)i+1} - \sum_{i=0}^{\infty} \frac{(K_t^*)^i}{i!} \frac{1}{(1-2\delta)i+1} l^{(1-2\delta)i+1}.$$

Since $-(i+1) < (1-2\delta)i + 1 < 1$,

$$\int_l^x \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}\eta^{1-2\delta}\right\} d\eta > x \sum_{i=0}^{\infty} \frac{\left(K_t^* x^{1-2\delta}\right)^i}{i!} + l \sum_{i=0}^{\infty} \frac{\left(K_t^* l^{1-2\delta}\right)^i}{(i+1)!}.$$

Then

$$\int_l^x \exp\left\{-\frac{2L(t)}{\sigma^2(1-2\delta)}\eta^{1-2\delta}\right\} d\eta > x \exp\left\{K_t^* x^{1-2\delta}\right\} + \frac{l^{2\delta}}{K_t^*}\left(\exp\left\{K_t^* l^{1-2\delta}\right\} - 1\right).$$

$\lim_{l\to 0^+} \frac{l^{2\delta}}{K_t^*} \exp\left\{K_t^* l^{1-2\delta}\right\} = \infty$ implies that $\boldsymbol{Q}((0,x],t) = \lim_{l\to 0^+} \boldsymbol{Q}([l,x],t) = \infty$. From Lemma 6.3 [Karlin and Taylor, 1981, Page 231], $\boldsymbol{Q}((0,x],t) = \infty$ implies $p(0) = \infty$. From Definition C.2 in [Karlin and Taylor, 1981, Page 230], 0 is always an unattainable boundary for $1/2 < \delta < 1$.

(3). This proof is similar to step (1) or (2).

(4). If $\delta = 1/2$, the model in consideration is the case of generalized CIR model. By similar calculation, it finishes the proof. □

Proof of Proposition 4.2.2. To apply Proposition 2.13 (*Yamada-Watanabe Theorem* 1971) Karatzas and Shreve [1998], the necessary step is to verify that all the requirements are met ($\forall\ x, y \geqslant 0$).

(1) First, by (4.2.2),

$|S(t,x) - S(t,y)| = |(L(t) - \alpha x) - (L(t) - \alpha y)| = \alpha|x-y|$, which implies that the drift term $S(t,x)$ satisfies Lipschitz condition.

(2) Second, the function σx^δ vanishes at 0 and satisfies the Hölder condition, which is

$$\left|\sigma x^\delta - \sigma y^\delta\right| \leqslant \sigma|x-y|^\delta.$$

Indeed, by choosing $h(u) = \sigma u^\delta$, $|\sigma(t,x) - \sigma(t,y)| = \sigma|x^\delta - y^\delta|$. Further, $|x+y|^\delta \leqslant |x|^\delta + |y|^\delta$. Without loss of generality, suppose $0 < y < x$. Replacing x by $x - y$,

$|x|^{\delta} - |y|^{\delta} \leqslant |x-y|^{\delta}$, which implies that $\left||x|^{\delta} - |y|^{\delta}\right| \leqslant |x-y|^{\delta} = \frac{1}{\sigma}h(|x-y|)$.

By Proposition 2.13 of (*Yamada-Watanabe Theorem*) Karatzas and Shreve [1998], the SDE (4.2.2) admits a non-negative strong and unique solution in $(0, \infty)$. This completes the proof of the first statement. Further, by Proposition 4.2.1, and Assumption 4.2, 0 is an unattainable boundary for (4.2.1). Let $\mathbb{R}_+ := [0, +\infty)$ and $\mathbb{R}_+^* := (0, +\infty)$, one has

- The coefficients $\sigma(x), \beta - \alpha x, L(t) - \alpha x$ are continuous, real-valued functions on $\mathbb{R}_+ \times \mathbb{R}_+^*$.
- In the proof of the first statement, it is established that $\sigma(x) = x^{\delta}$ satisfies $|\sigma(x) - \sigma(y)| \leqslant h(|x-y|) = \sigma|x-y|^{\delta}$, where $h : [0, +\infty) \to [0, +\infty)$ is a strictly increasing function with $h(0) = 0$ and $\int_0^{\varepsilon} h^{-2}(u)du = \infty;\ \forall\, \varepsilon > 0$.
- Assumption 4.1 implies that $r_0 \leqslant X_0$ a.s.
- Assumption 4.2 implies $\beta - \alpha x \leqslant L(t) - \alpha x,\ \forall\ 0 \leqslant t < \infty,\ x \in \mathbb{R}$.
- Both $\beta - \alpha x$ and $L(t) - \alpha x$ satisfy the Lipschitz condition in x.

Then, from Proposition B.2, that $0 < r_t \leqslant X(t)$ a.s. This completes the proof. □

Proof of Proposition 4.2.3. Let $G(t, X(t)) = e^{\alpha t}X(t)$, by Itô's Lemma,

$$X(t) = e^{-\alpha t}X_0 + e^{-\alpha t}\int_0^t e^{\alpha s}L(s)ds + \sigma e^{-\alpha t}\int_0^t e^{\alpha s}X^{\delta}(t)dB_s. \tag{C.15}$$

This completes the proof. □

Proof of Proposition 4.2.4. For the given SDE (4.2.1), from Proposition 2.1 in Andersen and Piterbarg [2007], 0 is an unattainable boundary for $\delta > 1/2$ and ∞ is an unattainable boundary for all values of $\delta > 0$. By Assumption 4.2, Proposition B.2, and Proposition 2.1 in Mishura et al. [2022], comparison of the solutions concludes that $\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{1-2\delta}\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[(r_t)^{1-2\delta}\right] < \infty$, $\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-2\delta}\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[(r_t)^{-2\delta}\right] < \infty$, $\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-1}\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[(r_t)^{-1}\right] < \infty$. □

Proof of Proposition 4.2.5. (1) Since Itô integrals are local martingales (see Theorem 13.2 in Klebaner [2005]), $\int_0^t X^\delta(t)dB_s$ is a local martingale. For $n \in \mathbb{N}$, define a stopping time $T_n(\omega)$ by $T_n(\omega) := \inf\{t \geqslant 0, X(t,\omega) \geqslant n\}$. Since $\{X(t), t \geqslant 0\}$ has almost surely continuous sample paths, $X(t \wedge T_n) \leqslant n$. So that $\int_0^{t\wedge T_n} X^\delta(t)dB_s$ is a martingale in t for any fixed n, then from (4.2.2), denote

$$X(t \wedge T_n) = X_0 + \int_0^{t\wedge T_n} (L(s) - \alpha X(s))\, ds + \sigma \int_0^{t\wedge T_n} X^\delta(t)dB_s. \tag{C.16}$$

Taking expectation both sides,

$$\mathbb{E}[X(t \wedge T_n)] = \mathbb{E}[X_0] + \mathbb{E}\left[\int_0^{t\wedge T_n} (L(s) - \alpha X(s))ds\right]. \tag{C.17}$$

Since $X(t)$ is non-negative, $\int_0^{t\wedge T_n(\omega)} X(s,\omega)ds$ is increasing and $\lim_{n\to\infty}\int_0^{t\wedge T_n(\omega)} X(s,\omega)ds = \int_0^t X(s,\omega)ds$ for all $\omega \in \Omega$, $\int_0^{t\wedge T_n} X(s)ds \xrightarrow[n\to\infty]{a.s.} \int_0^t X(s)ds$. Therefore, by monotone convergence theorem, $\int_0^{t\wedge T_n} X(s)ds \xrightarrow[n\to\infty]{L^1} \int_0^t X(s)ds$, which implies that

$$\lim_{n\to\infty}\mathbb{E}\left[\int_0^{t\wedge T_n} X(s)ds\right] = \mathbb{E}\left[\int_0^t X(s)ds\right]. \tag{C.18}$$

Similarly, $\int_0^{t\wedge T_n} L(s)ds \xrightarrow[n\to\infty]{a.s.} \int_0^t L(s)ds$, so, $\lim_{n\to\infty}\mathbb{E}\left[X(t \wedge T_n)\right] = \mathbb{E}[X_0] + \mathbb{E}[\int_0^t (L(s) - \alpha X(s))ds]$. Since $X(t \wedge T_n) \xrightarrow[n\to\infty]{a.s.} X(t)$, by using Fatou's Lemma, one has

$$\mathbb{E}[X(t)] = \mathbb{E}[\liminf_{n\to\infty} X(t\wedge T_n)] \leqslant \liminf_{n\to\infty}\mathbb{E}[X(t\wedge T_n)] = \mathbb{E}[X_0] + \int_0^t L(s)ds - \alpha\mathbb{E}\left[\int_0^t X(s)ds\right].$$

Since $\alpha > 0$, $\mathbb{E}\left[\int_0^t X(s)ds\right] > 0$, $\mathbb{E}[X(t)] < \mathbb{E}[X_0] + \int_0^t L(s)ds$. Let $\sum_{i=1}^p |\mu_i| \leqslant K_\mu$ and $\|\varphi(t)\| \leqslant K_\varphi$, for some positive constants K_μ, K_φ, then,

$$0 < L(t) \leqslant K_\varphi K_\mu. \tag{C.19}$$

$$\mathbb{E}[X(t)] < \mathbb{E}[X_0] + K_\mu K_\varphi t. \tag{C.20}$$

Further, let $G(x) = x^2$, by Itô's lemma,

$$X^2(t) = X_0^2 + \int_0^t \left(2X(s)(L(s) - \alpha X(s)) + \sigma^2 X^{2\delta}(t)\right) ds + \sigma \int_0^t X^\delta(s) dB_s. \tag{C.21}$$

For $n \in \mathbb{N}$, define a stopping time $T_n(\omega)$ by $T_n(\omega) := \inf\{t \geqslant 0, X(t)^2(\omega) \geqslant n\}$. let $T_n = T_n(\omega)$. Since $\{X(t)^2, t \geqslant 0\}$ has almost surely continuous sample paths, it holds that $X(t \wedge T_n)^2 \leqslant n$. From the inequality $X^\delta(t) \leqslant \max\{1, X^2(t)\}$, one obtains that $\int_0^{t\wedge T_n} X^\delta(t) dB_s$ is a martingale in t for any fixed n, then (C.21) implies that

$$X^2(t \wedge T_n) = X_0^2 + \int_0^{t\wedge T_n} \left(2X(s)(L(s) - \alpha X(s)) + \sigma^2 X^{2\delta}(t)\right) ds + \sigma \int_0^{t\wedge T_n} X^\delta(s) dB_s. \tag{C.22}$$

Taking expectation both sides,

$$\mathbb{E}\left[X^2(t \wedge T_n)\right] = \mathbb{E}\left[X_0^2\right] + \mathbb{E}\left[\int_0^{t\wedge T_n} \left(2X(s)(L(s) - \alpha X(s)) + \sigma^2 X^{2\delta}(t)\right) ds\right]. \tag{C.23}$$

Since $X(t), L(t)$ are non-negative, $\int_0^{t\wedge T_n(\omega)} X(s,\omega)L(s)ds$ is increasing and $\lim_{n\to\infty} \int_0^{t\wedge T_n(\omega)} X(s,\omega)L(s)ds = \int_0^t X(s,\omega)L(s)ds$ for all $\omega \in \Omega$, $\int_0^{t\wedge T_n} X(s)L(s)ds \xrightarrow[n\to\infty]{a.s.} \int_0^t X(s)L(s)ds$. Therefore, by monotone convergence theorem, $\int_0^{t\wedge T_n} X(s)L(s)ds \xrightarrow[n\to\infty]{L^1} \int_0^t X(s)L(s)ds$, which implies that

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^{t\wedge T_n} X(s)L(s)ds\right] = \mathbb{E}\left[\int_0^t X(s)L(s)ds\right]. \tag{C.24}$$

Similarly, from $X(t)^2 > 0,\ X^{2\delta}(t) > 0$,

$$\lim_{n\to\infty} \mathbb{E}\left[\int_0^{t\wedge T_n} X^2(s)ds\right] = \mathbb{E}\left[\int_0^t X^2(s)ds\right], \quad \lim_{n\to\infty} \mathbb{E}\left[\int_0^{t\wedge T_n} X^{2\delta}(t)ds\right] = \mathbb{E}\left[\int_0^t X^{2\delta}(t)ds\right]. \tag{C.25}$$

Then, $\lim_{n\to\infty} \mathbb{E}\left[X(t\wedge T_n)^2\right] = \mathbb{E}\left[X_0^2\right] + \mathbb{E}\left[\int_0^t \left(2X(s)(L(s) - \alpha X(s)) + \sigma^2 X^{2\delta}(t)\right) ds\right]$. Since $X(t \wedge T_n)^2 \xrightarrow[n\to\infty]{a.s.} X(t)^2$, by using Fatou's Lemma,

$$\begin{aligned}\mathbb{E}[X(t)^2] &= \mathbb{E}[\liminf_{n\to\infty} X(t\wedge T_n)^2] \leqslant \liminf_{n\to\infty} \mathbb{E}[X(t\wedge T_n)^2] \\ &= \mathbb{E}\left[X_0^2\right] + \mathbb{E}\left[\int_0^t \left(2X(s)(L(s) - \alpha X(s)) + \sigma^2 X^{2\delta}(t)\right) ds\right].\end{aligned}$$

Then, $\mathbb{E}[X(t)^2] < \mathbb{E}\left[X_0^2\right] + 2\int_0^t \mathbb{E}\left[X(s)\right] L(s)ds + \sigma^2 \int_0^t \mathbb{E}\left[X^{2\delta}(t)\right] ds$. From the inequality $X^{2\delta}(t) \leqslant \max\{1, X^2(t)\}$, and $\mathbb{E}\left[X^{2\delta}(t)\right] \leqslant \mathbb{E}\left[\max\{1, X^2(t)\}\right] \leqslant \mathbb{E}\left[X^2(t) + 1\right]$,
$\mathbb{E}[X(t)^2] < \mathbb{E}\left[X_0^2\right] + 2\int_0^t \mathbb{E}\left[X(s)\right] L(s)ds + \sigma^2 t + \sigma^2 \int_0^t \mathbb{E}\left[X^2(t)\right] ds$. From (C.19) and (C.20),

$$\mathbb{E}[X(t)^2] < \mathbb{E}\left[X_0^2\right] + 2K_\varphi K_\mu \left(\mathbb{E}[X_0]t + \frac{1}{2}K_\varphi K_\mu t^2\right) + \sigma^2 t + \sigma^2 \int_0^t \mathbb{E}\left[X^2(t)\right] ds.$$

By Grönwall inequality,

$$\begin{aligned} \mathbb{E}[X(t)^2] &< \mathbb{E}\left[X_0^2\right] + 2K_\varphi K_\mu \left(\mathbb{E}[X_0]t + \tfrac{1}{2}K_\varphi K_\mu t^2\right) + \sigma^2 t \\ &+\sigma^2 \int_0^t \left(\mathbb{E}\left[X_0^2\right] + 2K_\varphi K_\mu \left(\mathbb{E}[X_0]s + \frac{1}{2}K_\varphi K_\mu s^2\right) + \sigma^2 s\right) e^{t-s} ds. \end{aligned} \tag{C.26}$$

Then, the quadratic variation $\left\langle \int_0^\cdot e^{-\alpha(t-s)}\sigma X^\delta(t)dB_s\right\rangle_t = \int_0^t e^{-2\alpha(t-s)}\sigma^2 X^{2\delta}(t)ds$, and by (C.26),

$$\begin{aligned} \mathbb{E}\left[\int_0^t e^{-2\alpha(t-s)}\sigma^2 X^{2\delta}(t)ds\right] < \int_0^t e^{-2\alpha(t-s)}\sigma^2 \mathbb{E}\left[X_0^2\right] + 2K_\varphi K_\mu \left(\mathbb{E}[X_0]s + \frac{1}{2}K_\varphi K_\mu s^2\right) + \sigma^2 s \\ + \sigma^2 \int_0^s \left(\mathbb{E}\left[X_0^2\right] + 2K_\varphi K_\mu \left(\mathbb{E}[X_0]u + \frac{1}{2}K_\varphi K_\mu s^2\right) + \sigma^2 u\right) e^{s-u} du\right) ds < \infty, \end{aligned}$$

which is

$$\mathbb{E}\left[\left\langle \int_0^\cdot e^{-\alpha(t-s)}\sigma X^\delta(t)dB_s\right\rangle_t\right] < +\infty. \tag{C.27}$$

Thus, the term $\displaystyle\int_0^t e^{-\alpha(t-s)}\sigma X^\delta(t)dB_s$ is a integrable martingale. Then, from (C.15),

$$\mathbb{E}[X(t)] = e^{-\alpha t}\mathbb{E}[X_0] + e^{-\alpha t}\int_0^t e^{\alpha s}L(s)ds = e^{-\alpha t}\mathbb{E}[X_0] + \sum_{i=1}^p \mu_i \int_0^t e^{-\alpha(t-s)}\varphi_i(s)ds. \tag{C.28}$$

This completes the proof of the first assertion.

(2) From Assumption 4.1 and (C.15),

$$\begin{aligned} \mathbb{E}[X^2(t)] = e^{-2\alpha t}\mathbb{E}\left[X_0^2\right] + e^{-2\alpha t}\left(\int_0^t e^{\alpha s}L(s)ds\right)^2 + \sigma^2 e^{-2\alpha t}\mathbb{E}\left[\left(\int_0^t e^{\alpha s}X^\delta(t)dB_s\right)^2\right] \\ +2e^{-2\alpha t}\mathbb{E}\left[X_0\right] \int_0^t e^{\alpha s}L(s)ds + 2\sigma e^{-2\alpha t}\int_0^t e^{\alpha s}L(s)ds\, \mathbb{E}\left[\int_0^t e^{\alpha s}X^\delta(t)dB_s\right] \\ +2e^{-\alpha t}\mathbb{E}[X_0]\, \sigma e^{-\alpha t}\mathbb{E}\left[\int_0^t e^{\alpha s}X^\delta(t)dB_s\right]. \end{aligned}$$

By Itô isometry and (C.28), $\mathbb{E}\left[\left(\int_0^t e^{\alpha s} X^{\delta}(t) dB_s\right)^2\right] = \int_0^t e^{2\alpha s}\mathbb{E}\left[X^{2\delta}(t)\right] ds$. From (C.27), $\int_0^t e^{\alpha s} X^{\delta}(t) dB_s$ is a martingale, then, $\mathbb{E}\left[\int_0^t e^{\alpha s} X^{\delta}(t) dB_s\right] = 0$. Finally,

$$\mathbb{E}[X^2(t)] = e^{-2\alpha t}\mathbb{E}\left[X_0^2\right] + e^{-2\alpha t}\left(\int_0^t e^{\alpha s} L(s) ds\right)^2 + \sigma^2 e^{-2\alpha t}\left(\int_0^t e^{2\alpha s}\mathbb{E}\left[X^{2\delta}(t)\right] ds\right)$$

$$+2e^{-2\alpha t}\mathbb{E}\left[X_0\right] \int_0^t e^{\alpha s} L(s) ds.$$

From (C.19) and together with Theorem 1.2.3 [Qin, 2016, page 11],

$$\mathbb{E}[X^2(t)] \leqslant e^{-2\alpha t}\mathbb{E}\left[X_0^2\right] + \sigma^2\left(\left(\frac{K_\mu K_\varphi}{\sigma}\right)^2 + 2\mathbb{E}\left[X_0\right] K_\mu K_\varphi \frac{1}{\sigma\alpha}\right)\left(2\alpha - \sigma^2\right)^{-1}\left(1 - e^{-\left(2\alpha-\sigma^2\right)t}\right).$$

Then, if $2\alpha > \sigma^2$, $\sup\limits_{t \geqslant 0}\mathbb{E}[X^2(t)] \leqslant \mathbb{E}\left[X_0^2\right] + \sigma^2\left(\left(\frac{K_\mu K_\varphi}{\sigma}\right)^2 + 2\mathbb{E}\left[X_0\right] K_\mu K_\varphi \frac{1}{\sigma\alpha}\right)\left(2\alpha - \sigma^2\right)^{-1} < \infty$. This completes the proof. □

Proof of Corollary 4.2.1. By (4.2.2), $S(\theta, t, X(t)) = L(t) - \alpha X(t)$, and $L(t) = \sum\limits_{i=1}^{p} \mu_i \varphi_i(t)$, which implies that

$$\left(\frac{S(\theta, t, X(t))}{\sigma X^{\delta}(t)}\right)^2 = \frac{L^2(t) - 2\alpha L(t) X(t) + X^2(t)}{\sigma^2 X^{2\delta}(t)}$$

$$= \frac{L^2(t)}{\sigma^2}(X(t))^{-2\delta} - \frac{2\alpha L(t)}{\sigma^2}(X(t))^{1-2\delta} + \frac{1}{\sigma^2}(X(t))^{2-2\delta}.$$

From (4.2.4) and (4.2.5), $\sup\limits_{t \geqslant 0}\mathbb{E}\left[\left(\frac{S(\theta, t, X(t))}{\sigma X^{\delta}(t)}\right)^2\right] < \infty$, and then, $\mathbb{E}\left[\int_0^T \left(\frac{S(\theta, t, X(t))}{\sigma X^{\delta}(t)}\right)^2 dt\right] < \infty$. Hence $\mathrm{P}\left(\int_0^T \left(\frac{S(\theta, t, X(t))}{\sigma X^{\delta}(t)}\right)^2 dt < \infty\right) = 1$, for all $0 \leqslant T < \infty$. This completes the proof. □

Proof of Proposition 4.3.1. From Assumption 2.2, all coefficients of CKLS are analytic on $[0, \infty) \times \mathbb{R}_+^*$. By the definition of the drift term $S(t, x)$, Assumption C.2 a) is satisfied. Next step is to define a function $V : (0, +\infty) \mapsto [1, +\infty)$ as $V(x) = 1 + x + |\log x|$ which is a Lyapunov function for the skeleton chain $\mathbb{X} = (X_k)_{k \in \mathbb{N}_0}$. Let $x \in C_m$, $\mathbb{P}_{0,1} V(x) = 1 + \mathbb{E}[X_1 | X_0 = x] + \mathbb{E}[|\log X_1| | X_0 = x]$. From (C.15) and Assumption 4.1, $\mathbb{E}[X_1 | X_0 = x] =$

$e^{-\alpha}x + e^{-\alpha}\int_0^1 e^{\alpha s}L(s)ds$. Further, let $Z(t, X(t)) = e^{\alpha t}\log X(t)$. By using Itô's lemma,

$$\log X(t) = e^{-\alpha t}\log X_0 + e^{-\alpha t}\int_0^t \left(\alpha e^{\alpha s}\log X(s) + e^{\alpha s}\frac{L(s)}{X(s)} - \alpha e^{\alpha s} - \frac{\sigma^2}{2}e^{\alpha s}(X(t))^{2\delta-2}\right)ds$$
$$+\sigma e^{-\alpha t}\int_0^t e^{\alpha s}X^{\delta-1}(t)dB_s.$$

Then,

$$\left|\log X(t)\right| = \left|e^{-\alpha t}\log X_0 + e^{-\alpha t}\int_0^t \left(\alpha e^{\alpha s}\log X(s) + e^{\alpha s}\frac{L(s)}{X(s)} - \alpha e^{\alpha s} - \frac{\sigma^2}{2}e^{\alpha s}(X(t))^{2\delta-2}\right)ds\right.$$
$$\left.+\sigma e^{-\alpha t}\int_0^t e^{\alpha s}X^{\delta-1}(t)dB_s\right|,$$

and since $X(t) > 0$ for $t \geqslant 0$,

$$\left|\log X(t)\right| \leqslant e^{-\alpha t}\left|\log X_0\right| + e^{-\alpha t}\int_0^t \left(\alpha e^{\alpha s}\left|\log X(s)\right| + e^{\alpha s}\frac{L(s)}{X(s)}\right)ds + \sigma e^{-\alpha t}\left|\int_0^t e^{\alpha s}X^{\delta-1}(t)dB_s\right|.$$

Further, for given $X_0 = x$,

$$|\log X(t)| \leqslant e^{-\alpha t}|\log x| + e^{-\alpha t}\int_0^t \left(\alpha e^{\alpha s}|\log X(s)| + e^{\alpha s}\frac{L(s)}{X(s)}\right)ds + \sigma e^{-\alpha t}\left|\int_0^t e^{\alpha s}X^{\delta-1}(t)dB_s\right|.$$

Since the function $|\log x| < \max\left\{x, \frac{1}{x}\right\}$, from Proposition 4.2.4,

$$\mathbb{E}[|\log X(t)|] \leqslant \sup_{t\geqslant 0}\left\{\mathbb{E}[X(t)] + \mathbb{E}\left[\frac{1}{X(t)}\right]\right\} < \infty.$$

Further, for the Itô integral term, from (4.2.5), $\sigma e^{-\alpha t}\left|\int_0^t e^{\alpha s}X^{\delta-1}(t)dB_s\right|$ is L^2-bounded. Finally, there exists a constant $c_1 > 0$, such that $\mathbb{P}_{0,1}V(x) \leqslant 1 + e^{-\alpha}x + e^{-\alpha}|\log x| + c_1$. It is obvious that for $x \in C_m$, $\mathbb{P}_{0,1}V(x)$ is bounded. For $x \in \mathbb{R}_+^*\backslash C_m$,

$$\mathbb{P}_{0,1}V(x) \leqslant V(x) - (1 - e^{-\alpha})x - (1 - e^{-\alpha})|\log x| + c_1 < V(x) - \varepsilon,$$

for some $m > 0, \varepsilon > 0$. This proves the Assumption C.2 b).

Now, the attention is turning to the key Assumption C.3. to specify a point x^* in $int(\mathbb{R}_+^*)$ of full weak Höermander dimension and attainable in a sense of deterministic control. Our candidate is $x^* = 1$. It is clear that $x^* \in \mathbb{R}_+^*$. From Definition C.5, the control systems $t \mapsto \phi(t)$ is related to Stratonovich drift. For the generalized CKLS model, from Definition C.4, the Stratonovich drift is as follows:

$$\tilde{S}(t,x) = S(t,x) - \frac{1}{2}\sigma x^{\delta}\frac{\partial}{\partial x}(\sigma x^{\delta}) = S(t,x) - \frac{1}{2}\sigma^2\delta x^{2\delta-1}.$$

Thus, for the given initial value $x \in \mathbb{R}^*$, it is needed to construct a function $\dot{h} \in L^2_{loc}$, which determines the paths of $\phi(s)$, satisfying the corresponding deterministic integral equation

$$\frac{d}{ds}\phi(s) = \tilde{S}(t,x) + \sigma(\phi(s))^{\delta}\dot{h}(s) \tag{C.29}$$

from the starting point $x = \phi(0)$ to $x^* = \lim_{t\to\infty}\phi(t)$. For the purpose of pushing the solution to (C.29) towards to the expected limit 1, a C^{∞}-function $\Phi^{(x,1)}$ is chosen with the following property

$$\begin{cases} \Phi^{(x,1)}(t) > 0 \; for \; all \; t \geqslant 0 \; and \; all \; x > 0 \\ \Phi^{(x,1)}(0) = x \\ \Phi^{(x,1)}(t) = 1, \; for \; all \; t \geqslant |x-1| + 1 \\ \left|\frac{d}{dt}\Phi^{(x,1)}\right| \leqslant 1 \; for \; all \; t \geqslant 0 \; and \; all \; x > 0. \end{cases}$$

Now, let $\phi(t) := \Phi(t)$ for all $t \geqslant 0$. Then, (C.29) determines the control function h, with

$$\dot{h}(s) = \frac{\frac{d}{ds}\phi(s) - \tilde{S}(t,x)}{\sigma(\phi(s))^{\delta}} = \frac{\frac{d}{ds}\Phi(s) - \tilde{S}(t,x)}{\sigma(\Phi(s))^{\delta}}, \quad t \geqslant 0.$$

By the construction of the function $\Phi(t)$, $\dot{h}(s) \in L^2_{loc}$, and the C^{∞}-function $\Phi^{(x,1)}$ has the limit 1. This implies that the point $x^* = 1 \in \mathbb{R}^*_+$ is attainable.

The following is to prove that x^* is of full weak Höermander dimension. From the previous discussion, $\tilde{S}(t,x) = S(t,x) - \frac{1}{2}\sigma^2\delta x^{2\delta-1} = L(t) - \alpha x - \frac{1}{2}\sigma^2\delta x^{2\delta-1}$. Consider the vector fields $\bar{U} \mapsto \mathbb{R}^{1+1}$,

$$V_0 = \begin{pmatrix} 1 \\ L(t) - \alpha x - \frac{1}{2}\sigma^2\delta x^{2\delta-1} \end{pmatrix}, \quad V_1 = \begin{pmatrix} 0 \\ \sigma x^{\delta} \end{pmatrix}, \quad [V_0,\; V_1] = \mathbb{J}_{V_1}V_0 - \mathbb{J}_{V_0}V_1,$$

where $[V_0,\; V_1]$ is the Lie bracket of V_0 and V_1, and $\mathbb{J}_{V_0}$, $\mathbb{J}_{V_1}$ are Jacobian matrices (Definition 2.32 in Holbach [2018]). In this generalized CKLS model,

$$\mathbb{J}_{V_1} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma\delta x^{\delta-1} \end{pmatrix}, \quad \mathbb{J}_{V_0} = \begin{pmatrix} 0 & 0 \\ 0 & -\alpha - \frac{1}{2}\sigma^2\delta(2\delta-1)x^{2\delta-2} \end{pmatrix}.$$

Then,

$$[V_0,\ V_1] = \begin{pmatrix} 0 \\ \sigma\delta x^{\delta-1}\left(L(t) - \alpha x - \frac{1}{2}\sigma^2\delta x^{2\delta-1}\right) \end{pmatrix} - \begin{pmatrix} 0 \\ \sigma x^{\delta}\left(-\alpha - \frac{1}{2}\sigma^2\delta(2\delta-1)x^{2\delta-2}\right) \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ \sigma\delta x^{\delta-1}\left(L(t) - \alpha x - \frac{1}{2}\sigma^2\delta x^{2\delta-1}\right) + \sigma x^{\delta}\left(\alpha + \frac{1}{2}\sigma^2\delta(2\delta-1)x^{2\delta-2}\right) \end{pmatrix}.$$

When $x^* = 1 \in int(\mathbb{R}^*_+)$, this leads to

$$V_0(t, x^*) = \begin{pmatrix} 1 \\ L(t) - \alpha - \frac{1}{2}\sigma^2\delta \end{pmatrix}, \quad V_1(t, x^*) = \begin{pmatrix} 0 \\ \sigma \end{pmatrix},$$

$$[V_0,\ V_1](t, x^*) = \begin{pmatrix} 0 \\ \sigma\delta\left(L(t) - \alpha - \frac{1}{2}\sigma^2\delta\right) + \sigma\left(\alpha + \frac{1}{2}\sigma^2\delta(2\delta-1)\right) \end{pmatrix}$$

$$= \sigma \begin{pmatrix} 0 \\ \delta L(t) + \alpha(1-\delta) + \frac{1}{2}\sigma^2\delta(\delta-1) \end{pmatrix} = \sigma \begin{pmatrix} 0 \\ \delta L(t) + (1-\delta)(\alpha - \frac{1}{2}\sigma^2\delta). \end{pmatrix}$$

From Assumption 4.2, $\frac{1}{2} < \delta < 1$ and $2\alpha - \sigma^2 > 0$, then $\delta L(t) + (1-\delta)(\alpha - \frac{1}{2}\sigma^2\delta) > 0$, which implies that V_1 and $[V_0, V_1]$ are not linearly independent at the point (t, x^*) for all $t \in [0, 1]$. In this case, dim $(\Delta_{\mathcal{L}^*_N}) = 1$ on $[0, 1] \times \{x^*\}$ for all $N \geqslant 1$. This proves that Assumption C.3 holds. From Theorem 2.2 in Höpfner et al. [2016], it concludes that

1. The grid chain $\mathbb{X} = (X_k)_{k\in\mathbb{N}_0}$ is positive Harris recurrent.
2. The path segment $X = (X_{k+s})_{k\in\mathbb{N}_0, 0<s<1}$ chain is positive Harris recurrent.

□

C.3 On the proposed estimators and their relative efficiency

Proof of Proposition 4.4.1. From (4.2.2), $\langle X\rangle_t = \sigma^2 \int_0^t X^{2\delta}(t)ds$. Then,

$$\left(\langle X\rangle_{t+h} - \langle X\rangle_t\right)\Big/h = \sigma^2 \left(\int_t^{t+h} X^{2\delta}(t)ds\right)\Big/h \xrightarrow[h\to 0]{a.s.} \sigma^2 X^{2\delta}(t). \tag{C.30}$$

Similarly, for some $0 < s < T$ and $s \neq t$, $\frac{\langle X\rangle_{t+h}-\langle X\rangle_t}{\langle X\rangle_{s+h}-\langle X\rangle_s} \xrightarrow[h\to 0]{a.s.} \left(\frac{X(t)}{X(s)}\right)^{2\delta}$, which implies that $\log\left(\left(\langle X\rangle_{t+h} - \langle X\rangle_t\right)\Big/\left(\langle X\rangle_{s+h} - \langle X\rangle_s\right)\right) \xrightarrow[h\to 0]{a.s.} 2\delta \log\left(X(t)/X(s)\right)$. Therefore

$$\delta = \lim_{h\to 0} \log\left(\frac{\langle X\rangle_{t+h} - \langle X\rangle_t}{\langle X\rangle_{s+h} - \langle X\rangle_s}\right)\left(2\log\left(X(t)/X(s)\right)\right)^{-1}. \tag{C.31}$$

Further, from (C.30), $\left(\langle X\rangle_{t+h} - \langle X\rangle_t\right)\Big/\left(hX^{2\delta}(t)\right) = \sigma^2 \left(\int_t^{t+h} X^{2\delta}(t)ds\right)\Big/\left(hX^{2\delta}(t)\right) \xrightarrow[h\to 0]{a.s.} \sigma^2$. This completes the proof. □

Proof of Proposition 4.4.2. Let $a = [a_1^\top, a_2]$ with a_1 a $p-$column vector, and a_2 a scalar. $aQ_{[0,T]}a^\top = (a_1^\top, a_2)\boldsymbol{Q}_{[0,T]}(a_1^\top, a_2)^\top$. Then

$$\begin{aligned} aQ_{[0,T]}a^\top &= a_1^\top \int_0^T \varphi^\top(t)\varphi(t)(X(t))^{-2\delta}dta_1 - 2a_2 \int_0^T \varphi(t)(X(t))^{1-2\delta}dta_1 \\ &+ a_2 \int_0^T (X(t))^{2-2\delta}dta_2, \end{aligned}$$

and then, $aQ_{[0,T]}a^\top = \int_0^T \left(a_1^\top\varphi^\top(t)(X(t))^{-\delta} - a_2(X(t))^{1-\delta}\right)^2 dt$. Thus, $a^\top\boldsymbol{Q}_{[0,T]}a = 0$ if and only if

$\mathrm{P}\left(\omega : a_1^\top\varphi^\top(t)(X(t))^{-\delta}(\omega) - a_2(X(t))^{1-\delta}(\omega) = 0,\ 0 \leqslant t \leqslant T\right) = 1$, which is equivalent to

$$\mathrm{P}\left(\omega : a_1^\top\varphi^\top(t) - a_2X(t,\omega) = 0,\ 0 \leqslant t \leqslant T\right) = 1. \tag{C.32}$$

From Proposition 4.2.5,

$$\mathrm{Var}(X(t)) = e^{-2\alpha t}\left(\mathbb{E}\left[X_0^2\right] - \mathbb{E}^2\left[X_0\right]\right) + \sigma^2 e^{-2\alpha t}\left(\int_0^t e^{2\alpha s}\mathbb{E}\left[X^{2\delta}(t)\right]ds\right) > 0,$$

which implies that $X(t)$ is not a constant. Thus, if $a_2 \neq 0$, $a_2^2\mathrm{Var}(X(t)) > 0$ for all $t \geqslant 0$. So, from Proposition 4.2.2, $\mathrm{P}\left(\omega : a_1^\top\varphi^\top(t) - a_2X(t,\omega) = 0,\ 0 \leqslant t \leqslant T\right) = 0$. This is a contradiction with (C.32). So, the assumption $a_2 \neq 0$ is not correct, which implies that $a_2 = 0$. From $a_1^\top\varphi^\top(t) - a_2X(t,\omega) = 0$ in (C.32), $a_1^\top\varphi^\top(t) = 0, \forall\ t \geqslant 0$. If $T \geqslant 1$,

$[0,1] \subset [0,T]$, by Assumption 4.2, $\{\varphi_1(t), \varphi_2(t), \ldots, \varphi_p(t)\}$ is linearly independent on $[0,1]$, this implies that $a_1^\top \varphi^\top(t) = 0$ if and only if $a_1^\top = \vec{0}_{1\times(p+1)}$. Hence, if $T \geqslant 1$, the matrix $\boldsymbol{Q}_{[0,T]}$ is a positive definite matrix. This completes the proof. □

Proof of the Proposition 4.4.3. The likelihood function of the SDE in (4.2.2) is given by

$$\begin{aligned}\mathcal{L}(\theta, X^T) &= d\mathrm{P}^{(\theta)}_{X^T}/dP_B\\ &= \exp\left(\frac{1}{\sigma^2}\int_0^T S(t,\theta,X(t))(X(t))^{-2\delta}dX(t) - \frac{1}{2\sigma^2}\int_0^T S^2(t,\theta,X(t))(X(t))^{-2\delta}dt\right).\end{aligned}$$

Then, the log-likelihood function is given

$$\log \mathcal{L}(\theta, X^T) = \frac{1}{\sigma^2}\int_0^T S(t,\theta,X(t))(X(t))^{-2\delta}dX(t) - \frac{1}{2\sigma^2}\int_0^T S^2(t,\theta,X(t))(X(t))^{-2\delta}dt.$$

This gives $\log \mathcal{L}(\theta, X^T) = \frac{1}{\sigma^2}\theta^\top R_{[0,T]} - \frac{1}{2\sigma^2}\theta^\top \boldsymbol{Q}_{[0,T]}\theta$. Therefore, the proof follows from classical optimization techniques. This completes the proof. □

Proof of Proposition 4.4.5. By Itô isometry,

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{T}}\int_0^T \frac{\boldsymbol{\varphi}(t)}{X^\delta(t)}dB_t\right\|^2\right] = \frac{1}{T}\int_0^T \|\boldsymbol{\varphi}(t)\|^2\mathbb{E}\left[\frac{1}{X^{2\delta}(t)}\right]dt.$$

By Assumption 2.2 and the relation (4.2.4), $\|\varphi(t)\|^2 \leqslant K_\varphi^2$, and

$$\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-2\delta}\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[r_t^{-2\delta}\right] < \infty,$$

which implies that

$$\mathbb{E}\left[\left\|\frac{1}{\sqrt{T}}\int_0^T \frac{\boldsymbol{\varphi}(t)}{X^{2\delta}(t)}dB_t\right\|^2\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-2\delta}\right]K_\varphi^2\frac{1}{T}\int_0^T dt = \sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{-2\delta}\right]K_\varphi^2.$$

Further, $\mathbb{E}\left[\left(\frac{1}{\sqrt{T}}\int_0^T (X(t))^{1-2\delta}dB_t\right)^2\right] = \mathbb{E}\left[\frac{1}{T}\int_0^T (X(t))^{2(1-2\delta)}dt\right]$. Note that, since $1/2 < \delta < 1$, $-2 < 2(1-2\delta) < 0$. Then, by Assumption 4.2, Proposition B.2, and Proposition 2.1 in Mishura et al. [2022], $\sup_{t\geqslant 0}\mathbb{E}\left[(X(t))^{2(1-2\delta)}\right] \leqslant \sup_{t\geqslant 0}\mathbb{E}\left[r_t^{2(1-2\delta)}\right] < \infty$. This completes the proof of part (1) and part (2). In addition, from Proposition 4.4.4, $\frac{1}{T}\boldsymbol{Q}_{[0,T]} \xrightarrow[T\to\infty]{a.s.} \boldsymbol{\Sigma}$, a positive definite matrix. Then, Part (1) follows from the martingale strong law of large numbers for diffusion processes along with Slutsky's theorem. Part (2) follows from Proposition 4.4.4 and the martingale central limit for diffusion processes along with Slutsky's theorem. Part **(3)** follows from Propositions 4.4.4 and Part (1). Finally, $\rho_T = \sigma T \boldsymbol{Q}^{-1}_{[0,T]} \frac{1}{\sqrt{T}} W_{[0,T]}$. Then, by combining Proposition 4.4.4 and the martingale central limit theorem for diffusion processes along with Slutsky's theorem, $\rho_T \xrightarrow[T\to\infty]{\mathrm{D}} \rho \sim \mathcal{N}_{p+1}\left(0, \sigma^2 \boldsymbol{\Sigma}^{-1}\right)$. This completes the proof. □

Proof of Theorem 4.4.1. For every $\theta \in \Theta$, and arbitrary bounded sequences $\boldsymbol{h} \in \mathbb{R}^{p+1}$, the log-likelihood ratio of the SDE (4.2.2) is $\log(Z_T(h)) = dP^{(\theta)}_{X^T,\boldsymbol{h}}/dP_{B,0}$. This yields the representation

$$\log(Z_T(h)) = \int_0^T \frac{S(t, \theta + \boldsymbol{h}/\sqrt{T}, X(t)) - S(t, \theta, X(t))}{\sigma X^{\delta}(t)} dB_t$$

$$- \int_0^T \frac{\left(S(t, \theta + \boldsymbol{h}/\sqrt{T}, X(t)) - S(t, \theta, X(t))\right)^2}{2\sigma^2 X^{2\delta}(t)} dt,$$

by using the fact that $B_t = \int_0^t \frac{1}{\sigma X^{\delta}(t)} dX(s)$ is a $\mathscr{F}_t$ measurable Brownian motion. Then,

$$\log(Z_T(\boldsymbol{h})) = \frac{1}{\sigma} \boldsymbol{h}^{\top} \frac{1}{\sqrt{T}} \int_0^T \frac{(\varphi(t), -X(t))^{\top}}{X^{\delta}(t)} dB_t$$

$$- \frac{1}{2\sigma^2} \boldsymbol{h}^{\top} \left(\frac{1}{T} \int_0^T \frac{(\varphi(t), -X(t))^{\top} (\varphi(t), -X(t))}{X^{2\delta}(t)} dt \right) \boldsymbol{h}.$$

Letting $\Delta_T(\theta_0, X^T) = \frac{1}{\sqrt{T}} W_{[0,T]}$ and $r_{_T}(\theta_0, \boldsymbol{h}, X^T) = \frac{1}{2\sigma^2} \boldsymbol{h}^{\top} \left(\frac{1}{T} \boldsymbol{Q}_{[0,T]} - \boldsymbol{\Sigma} \right) \boldsymbol{h}$, (4.4.1) and (4.4.2) give

$$\log(Z_T(\boldsymbol{h})) = \frac{1}{\sigma} \boldsymbol{h}^{\top} \Delta_T(\theta_0, X^T) - \frac{1}{2\sigma^2} \boldsymbol{h}^{\top} \boldsymbol{\Sigma} \boldsymbol{h} - r_{_T}(\theta_0, \boldsymbol{h}, X^T).$$

The proof follows from Proposition 4.4.4 and Proposition 4.4.5. This completes the proof. □

Proof of Proposition 4.4.6. From (4.4.5), $\varsigma_T = G_{[0,T]} M \rho_T + \sqrt{T} G_{[0,T]} (\boldsymbol{M\theta} - r)$. Then, $(\rho_T, \varrho_T, \varsigma_T)' = \left(\boldsymbol{I}_{p+1}, \boldsymbol{I}_{p+1} - M' G'_{[0,T]}, M' G'_{[0,T]}\right)' \rho_T + \left(\boldsymbol{0}', -r_0' G'_{[0,T]}, r_0' G'_{[0,T]}\right)'$. By (4.4.6) and (4.4.7), $\left(\boldsymbol{I}_{p+1}, \boldsymbol{I}_{p+1} - M' G'_{[0,T]}, M' G'_{[0,T]}\right)' \xrightarrow[T\to\infty]{\mathrm{P}} \left(\boldsymbol{I}_{p+1}, \boldsymbol{I}_{p+1} - M' G^{*'}, M' G^{*'}\right)'$, and $\left(\boldsymbol{0}, -r_0' G'_{[0,T]}, r_0' G'_{[0,T]}\right)' \xrightarrow[T\to\infty]{\mathrm{P}} \left(\boldsymbol{0}, -r_0' G^{*'} r_0' G^{*'}\right)'$. Then, by Proposition 4.4.5 and Slutsky's

Theorem,
$\left(\rho_T', \varrho_T', \varsigma_T'\right)' \xrightarrow[T\to\infty]{D} \left(\boldsymbol{I}_{p+1}, \boldsymbol{I}_{p+1} - M'G^{*'}, M'G^{*'}\right)' \rho + \left(\boldsymbol{0}', -r_0'G^{*'}, r_0'G^{*'}\right)' = (\rho', \varrho', \varsigma')'$.
Then, the proof follows from some properties of multivariate normal distribution and algebraic computations. □

Proof of Proposition 4.5.1. From Proposition 4.4.6,

$$\varsigma_T \xrightarrow[T\to\infty]{D} \varsigma \sim \mathcal{N}_{p+1}(G^* r_0, \sigma^2 G^* M \boldsymbol{\Sigma}^{-1}).$$

Further, by Proposition 4.4.4,

$$\hat{\Gamma} = \frac{1}{\hat{\sigma}^2} M^\top \left(MTQ^{-1}M^\top\right)^{-1} M \xrightarrow[T\to\infty]{\mathrm{P}} \Gamma = \frac{1}{\sigma^2} M^\top \left(M\boldsymbol{\Sigma}^{-1}M^\top\right)^{-1} M.$$

Therefore, by Slutsky's Theorem, $\psi_T = \varsigma_T^\top \hat{\Gamma} \varsigma_T \xrightarrow[T\to\infty]{\mathrm{D}} \psi = \varsigma^\top \Gamma \varsigma$. Further, the proof follows from Theorem 5.1.3 in Mathai and Provost [1992], which is similar to the proof of Proposition 2.4.9. This completes the proof. □

Proof of Proposition 4.6.1. From Proposition 4.4.6,

$$\begin{aligned} \mathrm{ADR}\left(\hat{\theta}_T, \theta, \Omega\right) &= \mathbb{E}\left[\mathrm{trace}(\rho^\top \Omega \rho)\right] = \mathrm{trace}\left(\Omega \mathbb{E}\left[\rho\rho^\top\right]\right) = \sigma^2 \mathrm{trace}(\Omega \boldsymbol{\Sigma}^{-1}), \\ \mathrm{ADR}\left(\tilde{\theta}_T, \theta, \Omega\right) &= \mathbb{E}\left[\mathrm{trace}(\varrho^\top \Omega \varrho)\right] = \mathrm{trace}\left(\Omega \mathrm{Var}(\varrho)\right) + \mathbb{E}\left[\varrho^\top\right] \Omega \mathbb{E}\left[\varrho\right]. \end{aligned}$$

Therefore, by combining Proposition 4.4.6 and Proposition 4.6.1, the stated result. □

Proof of Proposition 4.6.2. By combining (4.5.2), Proposition 4.4.6, along with the fact that γ is continuous real-valued function on $(0, +\infty)$,

$$\begin{aligned} \mathrm{ADR}\left(\hat{\theta}^s, \theta, \Omega\right) = {} & \mathbb{E}\left[\mathrm{trace}\left(\varrho^\top \Omega \varrho\right)\right] + 2\mathbb{E}\left[\gamma\left(\|\varsigma\|_\Gamma^2\right) \mathrm{trace}\left(\varrho^\top \Omega \varsigma\right)\right] \\ & + \mathbb{E}\left[\gamma^2\left(\|\varsigma\|_\Gamma^2\right) \mathrm{trace}\left(\varsigma^\top \Omega \varsigma\right)\right]. \end{aligned}$$

Then, the proof follows from Theorem 3.1 in Nkurunziza [2012] along with some algebraic computations. This completes the proof. □

Proof of Proposition 4.6.3. Let $\gamma(x) = 1 - \frac{q-2}{x}, x > 0$. The proof follows by combining

Proposition 4.6.2 and the identity $\mathbb{E}\left[\chi_{q+2}^{-2}(\Delta)\right] - \mathbb{E}\left[\chi_{q+4}^{-2}(\Delta)\right] = 2\mathbb{E}\left[\chi_{q+4}^{-4}(\Delta)\right]$, along with some algebraic computations. $\square$

Bibliography

O. O. Aalen and H. K. Gjessing. Survival models based on the Ornstein-Uhlenbeck process. *Lifetime data analysis*, 10(4):407–423, 2004.

M. B. Alaya and A. Kebaier. Parameter estimation for the square-root diffusions: Ergodic and nonergodic cases. *Stochastic Models*, 28(4):609 – 634, 2012.

M. B. Alaya and A. Kebaier. Asymptotic behavior of the maximum likelihood estimator for ergodic and nonergodic square-root diffusions. *Stochastic Analysis and Applications*, 31(4):552 – 573, 2013.

L. B. Andersen and V. V. Piterbarg. Moment explosions in stochastic volatility models. *Finance and Stochastics*, 11(1):29–50, 2007.

A. Behme and A. Sideris. Markov-modulated generalized Ornstein-Uhlenbeck processes and an application in risk theory. *Bernoulli*, 28(2):1309–1339, 2022.

K. Ben Nowman and G. Sorwar. Implied option prices from the continuous time CKLS interest rate model: an application to the UK. *Applied Financial Economics*, 13(3): 191–197, 2003.

P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, 2013.

Y. Cai and S. Wang. Central limit theorem and moderate deviation principle for CKLS model with small random perturbation. *Statistics & Probability Letters*, 98:6–11, 03 2015. doi: 10.1016/j.spl.2014.11.017.

J. Casassus and P. Collin-Dufresne. Stochastic convenience yield implied from commodity futures and interest rates. *The Journal of Finance*, 60(5):2283–2331, 2005.

K. C. Chan, G. A. Karolyi, F. A. Longstaff, and A. B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *The journal of finance*, 47(3): 1209–1227, 1992.

F. Chen, R. Mamon, and S. Nkurunziza. Inference for a change-point problem under a generalised Ornstein–Uhlenbeck setting. *Annals of the Institute of Statistical Mathematics*, 70(4):807–853, 2018.

F. Chen, R. Mamon, and S. Nkurunziza. Inference for a change-point problem under an OU setting with unequal and unknown volatilities. *Canadian Journal of Statistics*, 48 (1):62–78, 2020.

R.-R. Chen and L. O. Scott. Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *The Journal of Fixed Income*, 3(3): 13–41, 1993.

S. Chen. *Modelling the dynamics of commodity prices for investment decisions under uncertainty*. PhD thesis, University of Waterloo, 2010.

A. Dassios, J. Jang, and H. Zhao. A generalised CIR process with externally-exciting and self-exciting jumps and its applications in insurance and finance. *Risks*, 7(4):103, 2019.

H. Dehling, B. Franke, and T. Kott. Drift estimation for a periodic mean reversion process. *Statistical Inference for Stochastic Processes*, 13(3):175–192, 2010.

H. Dehling, B. Franke, T. Kott, and R. Klperger. Change point testing for the drift parameters of a periodic mean reversion process. *Statistical Inference for Stochastic Processes*, 17:1–18, 2014.

O. Dehtiar, Y. Mishura, and K. Ralchenko. Two methods of estimation of the drift parameters of the Cox–Ingersoll–Ross process: Continuous observations. *Communications in Statistics-Theory and Methods*, pages 1–16, 2021.

A. K. Dixit and R. S. Pindyck. *Investment under uncertainty*. Princeton University Press, 1994.

A. Einstein. Investigations on the theory of the brownian movement. *Annals of Physic*, 1905.

B. Favetto. Parameter estimation by contrast minimization for noisy observations of a diffusion process. *Statistics*, 48(6):1344–1370, 2014.

X. Feng and D. Xie. Bayesian estimation of CIR model. *Journal of data science*, 10(2): 271–280, 2012.

H. Geman. *Commodities and commodity derivatives: modeling and pricing for agriculturals, metals and energy*. John Wiley & Sons, 2009.

S. Holbach. *Recurrence and parameter estimation for degenerate diffusions with internal variables and randomly perturbed time-inhomogeneous deterministic input*. PhD thesis, Universitätsbibliothek Mainz, 2018.

R. Höpfner, E. Löcherbach, and M. Thieullen. Ergodicity and limit theorems for degenerate diffusions with time periodic drift. Application to a stochastic Hodgkin-Huxley model. *ESAIM: Probability and Statistics*, 20:527–554, 2016.

Y. Hu, G. Lan, and C. Zhang. The explicit solution and precise distribution of CKLS model under Girsanov transform. *International Journal of Statistics and Probability*, 4, 01 2015. doi: 10.5539/ijsp.v4n1p68.

O. Ibe. *Markov processes for stochastic modeling*. Newnes, 2013.

K. Itô. *On stochastic differential equations*, volume 4. American Mathematical Society, 1951.

J. John C. Cox, Jonathan E. Ingersoll and S. A. Ross. An international general equilibrium model of asset prices. *Econometrica*, 53(2):363–384, 1985a.

J. John C. Cox, Jonathan E. Ingersoll and S. A. Ross. A theory of the term structure of interest rates. *Econometrica*, 53(2):385–408, 1985b.

I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer, New York, NY, second edition, 1998.

S. Karlin and H. E. Taylor. *A Second Course in Stochastic Processes*. Academic Press,New York, NY, 1981.

B. P. Kellerhals. *Financial pricing models in continuous time and Kalman filtering*, volume 506. Springer Science & Business Media, 2013.

C. Khor, A. H. Pooi, and K. H. Ng. Bond option pricing under the CKLS model. *In: Regional Conference on Applied and Engineering Mathematics (2nd). Proceedings, 30 - 31 May 2012, Penang*, 2012.

F. C. Klebaner. *Introduction To Stochastic Calculus With Applications*. World Scientific Publishing Company, 2005.

P. E. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag Berlin Heidelberg, 1999.

V. Koroliuk, D. Koroliouk, and S. Dovgyi. Diffusion process with evolution and its parameter estimation. *Cybernetics and Systems Analysis*, 56:732–738, 2020.

K. Kubilius and A. Medžiūnas. Positive solutions of the fractional SDEs with non-Lipschitz diffusion coefficient. *Mathematics*, 9(1):18, 2020.

K. Kubilius, Y. Mishura, K. Ralchenko, K. Kubilius, Y. Mishura, and K. Ralchenko. Drift parameter estimation in diffusion and fractional diffusion models. *Parameter Estimation in Fractional Diffusion Models*, pages 161–267, 2017.

Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer-Verlag, London, 2004.

T. C. Langetieg. A multivariate model of the term structure. *The Journal of Finance*, 35 (1):71–97, 1980.

P. Lansky and L. Sacerdote. The Ornstein-Uhlenbeck neuronal model with signal-dependent noise. *Physics Letters A*, 285(3-4):132–140, 2001.

A. Le Breton. *On continuous and discrete sampling for parameter estimation in diffusion type processes*, pages 124–144. Springer Berlin Heidelberg, Berlin, Heidelberg, 1976.

C. Leuwattanachotinan. *Model fitting of a two-factor arbitrage-free model for the term structure of interest rates using Markov chain Monte Carlo*. PhD thesis, Heriot-Watt University, 2011.

R. S. Liptser and A. N. Shiryaev. *Statistics of random process I: General Theory*. Springer, Berlin, 2001.

Z. Liu, Z. Peng, and Z. Wang. Exponential Ornstein-Uhlenbeck process and its application in value of investment project. *Journal of Changsha university of electric power (natural science)*, 21(4):65, 2006.

Y. Lyu and S. Nkurunziza. Estimation and testing in generalized CIR model. *Annals of Applied Probability (Under review)*, 2023a.

Y. Lyu and S. Nkurunziza. Inference methods in time-varying linear diffusion processes. *Bernoulli (Under review)*, 2023b.

Y. Lyu and S. Nkurunziza. Inference in generalized exponential O-U processes. *Statistical Inference for Stochastic Processes*, 26(3):581–618, 2023c.

Y. Lyu and S. Nkurunziza. Inference in generalized exponential O-U processes with change-point. *Statistical Inference for Stochastic Processes*, 2023d. doi: 10.1007/s11203-023-09293-z.

A. P. B. Maboulou and H. P. Mashele. Credit derivative valuation and parameter estimation for multi-factor affine CIR-type hazard rate model. *Journal of Mathematical Finance*, 05(3):273–285, 2015.

Y. Maghsoodi. Solution of the extended Cox, Ingersoll and Ross model of the term structure and analytical bond option valuation. *University of Southampton preprint O.R. 54A*, 1993.

Y. Maghsoodi. Solution of the extended CIR term structure and bond option valuation. *Mathematical Finance*, 6(1):89–109, 1996.

J. Masoliver and J. Perelló. Multiple time scales and the exponential Ornstein–Uhlenbeck stochastic volatility model. *Quantitative Finance*, 6(5):423 – 433, 2006.

A. M. Mathai and S. B. Provost. *Quadratic forms in random variables: Theory and applications*. Marcel Dekker Inc, New York, 1992.

Y. Mishura, K. Ralchenko, and O. Dehtiar. Parameter estimation in CKLS model by continuous observations. *Statistics & Probability Letters*, 184:109391, 2022.

J. P. P. Monsalve and F. H. M. Sanchez. Parameter estimation in mean reversion processes with periodic functional tendency. *arXiv preprint arXiv:1710.11534*, 2017. doi: https://doi.org/10.48550/arXiv.1710.11534.

J. N. Nielsen, H. Madsen, and P. C. Young. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control*, 24:83–94, 2000.

S. Nkurunziza. The risk of pretest and shrinkage estimators. *Statistics*, 46(3):305 – 312, 2012.

S. Nkurunziza. On combining estimation problems under quadratic loss: A generalization. *arXiv: Methodology*, 2015.

S. Nkurunziza. Estimation and testing in multivariate generalized Ornstein-Uhlenbeck processes with change-points. *Sankhya A*, pages 1–50, 2021.

S. Nkurunziza and S. E. Ahmed. Shrinkage drift parameter estimation for multi-factor Ornstein–Uhlenbeck processes. *Applied Stochastic Models in Business and Industry*, 26:103–124, 2010.

S. Nkurunziza and K. Fu. Improved inference in generalized mean-reverting processes with multiple change-points. *Electronic Journal of Statistics*, 13(1):1400–1442, 2019.

S. Nkurunziza and L. Shen. Inference in a multivariate generalized mean-reverting process with a change-point. *Statistical Inference for Stochastic Processes*, 23(1): 199–226, 2019.

S. Nkurunziza and P. P. Zhang. Estimation and testing in generalized mean-reverting processes with change-point. *Statistical Inference for Stochastic Processes*, 21(1): 191–215, 2018.

P. Nowak and M. Romaniuk. Application of the one-factor CIR interest rate model to catastrophe bond pricing under uncertainty. *Journal of Automation, Mobile Robotics and Intelligent Systems*, pages 19–27, 2014.

B. Oksendal. *Stochastic differential equations: an introduction with applications*. Springer Science & Business Media, 2013.

Q. Peng and H. Schellhorn. On the distribution of extended CIR model. *Statistics & Probability Letters*, 142:23–29, 2018.

J. Perelló, R. Sircar, and J. Masoliver. Option pricing under stochastic volatility: The exponential Ornstein-Uhlenbeck model. *Econometric Modeling: Derivatives e-Journal*, 6, 2008.

Y. Qin. *Integral and discrete inequalities and their applications : Volume I: Linear inequalities*. Springer, 2016.

D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. Springer, Berlin, Heidelberg, third edition, 1999.

R. V. Rohlfs, P. Harrigan, and R. Nielsen. Modeling gene expression evolution with an extended Ornstein-Uhlenbeck process accounting for within-species variation. *Scandinavian Journal of Statistics*, 37(2):200–220, 2010.

A. K. M. E. Saleh. *Theory of preliminary test and Stein-type estimation with applications*. John Wiley & Sons, lnc., Hoboken, New Jersey, 2006.

F. H. M. Sánchez and V. M. Gallego. Parameter estimation in mean reversion processes with deterministic long-term trend. *Journal of Probability and Statistics*, 2016, 2016.

E. S. Schwartz. The stochastic behavior of commodity prices: Implications for valuation and hedging. *The Journal of finance*, 52(3):923–973, 1997.

G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6(2):461–464, 1978.

S. E. Shreve. *Stochastic Calculus for Finance II: Continuous-Time Models*. New York: Springer, 2004.

W. F. Stout. *Almost sure convergence*. Academic Press, New York, 1974.

J. Tong and Z. Zhang. Exponential ergodicity of CIR interest rate model with random switching. *Stochastics and Dynamics*, 17(05):1750037, 2017.

J. Tong, Q. Meng, Z. Zhang, and Y. K. Lu. A note on ergodicity for CIR model with markov switching. *Communications in Statistics - Simulation and Computation*, 50 (5):1445 – 1458, 2021.

O. Vasicek. An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188, 1977.

C. A. M. Vega. Calibration of the exponential Ornstein–Uhlenbeck process when spot prices are visible through the maximum log-likelihood method. example with gold prices. *Advances in Differential Equations*, 2018(1):1–14, 2018.

S. I. Vrieze. Model selection and psychological theory: a discussion of the differences between the akaike information criterion (aic) and the bayesian information criterion (bic). *Psychological methods*, 17(2):228–243, 2012.

C. Wei. Parameter estimation for Chan-Karoli-Longstaff-Saunders model driven by small Lévy noises from discrete observations. *International Journal of Dynamical Systems and Differential Equations*, 10(4):373–382, 2020.

K. C. Ying and P. A. Hin. Prediction of interest rate using CKLS model with stochastic parameters. In *AIP Conference Proceedings*, pages 467–472, 2014.

Z. Zhang, J. Cao, J. Tong, and E. Zhu. Ergodicity of CIR type sdes driven by stable processes with random switching. *Stochastics*, 92(5):761 – 784, 2019.

Z. Zhang, T. Zhou, X. Jin, and J. Tong. Convergence of the euler-maruyama method for CIR model with markovian switching. *Math. Comput. Simul.*, 177:192–210, 2020.

Z. Zíková and B. Stehlíková. Convergence model of interest rates of CKLS type. *Kybernetika*, 48(3):567–586, 2012.

Vita Auctoris

Yunhong Lyu was born in Hebei China. She earned her Master's degree in Applied Mathematics from the University of Science and Technology Beijing, specializing in modeling the temporal dynamics of disease transmission. Afterward, she served as an assistant professor at Xingtai University in Hebei Province. Currently, she is pursuing her Ph.D. in Statistics at the University of Windsor under the supervision of Dr. Nkurunziza and is expected to graduate in the fall of 2023