ABSTRACT

Title of dissertation: MINIMUM DISPARITY ESTIMATOR IN CONTINUOUS TIME STOCHASTIC VOLATILITY MODEL

Ziliang Li
Doctor of Philosophy, 2010

Dissertation directed by: Professor Eric V. Slud
Department of Mathematics

In the study of finance, likelihood based or moment based methods are frequently used to estimate parameters for various kinds of models given the sampled return data. While the former method is not robust, the latter one suffers from loss of efficiency and high noise-to-signal ratio in the data. In this paper, we investigate the ergodic behavior of the bivariate series described by the Barndorff-Nielsen and Shephard (BN-S) stochastic volatility model. In particular, we study its β-mixing property and the differentiability of its stationary distribution. A robust and efficient estimation scheme for continuous models called the Negative Exponential Disparity Estimator (NEDE) is studied. We apply this method and the classical Method of Moments (MOM) to the BN-S model. Asymptotic properties of the NEDE and the MOM estimator are proved, implementation details are provided.

MINIMUM DISPARITY ESTIMATOR IN CONTINUOUS TIME STOCHASTIC VOLATILITY MODEL

by

Ziliang Li

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2010

Advisory Committee:
Dr. Eric V. Slud, Chair/Advisor
Dr. Sandra Cerrai
Dr. Benjamin Kedem
Dr. Paul J. Smith
Dr. Armand Makowski

Preface

It always surprises me how long I have been in this quest for knowledge and pursue of a higher degree from the first day I arrived at the campus. As I am finishing up my thesis with this preface, six years time filled with joy, upset, surprise and, to be frank, boredom feels like flying away in a blink of an eye. I think I was born with keen curiosity and love to explore. In the past six years, I jumped onto backyards of several disciplines before finally settled down in working on a thesis which binds together elements from financial mathematics, pure stochastic process and rigorous statistical inference. Although it was not as exhilarating as touring a city where various cultures meet and I had to compromise occasionally, I successfully carried through this task and made a firm step forward into extending an old algorithm to greater generality and applicability. The invaluable education experience I had from this investigation substantiates my belief in doing cross-disciplinary researches and I shall continue on this track, seeking diversity as a curious traveler.

Ziliang Li

University of Maryland

August 2010

Dedication

This thesis is dedicated to my parents

Li, Baikeng and Deng, Yifang

Acknowledgments

I would like to thank many people who offered their help and kindness to me during my study and stay in this campus.

I owe my deep gratitude to my advisor Dr. Eric Slud, because of whom my thesis is completed with a standard higher than what I can expect. Thanks to his generous, selfless and continuous support, I was able to stick to my plan and accomplish a multidisciplinary study which I am long for. I will always remember those principles which I learnt from him on conducting research with rigor and integrity.

I would like to thank my committee members, Drs. Benjamin Kedem, Paul Smith, Sandra Cerrai and Armand Makowski for their constructive critics and suggestions during my oral examination. In particular, I would like to thank Dr. Benjamin Kedem and Dr. Paul Smith again for their valuable advises and guidance in the first two years of my graduate study. I also like to thank Dr. Dilip Madan and Dr. Michael Fu who organize the Mathematical Finance RIT which broadens my scope of statistical applications.

I want to express my true thankfulness to my girl friend Xuan Li, who provides endless support to my study, raises my spirit and encourages me to keep moving forward.

And I will never forget the great friends I have for the past six years. Tinghui Yu, Yabing Mai, Shihua Wen, Bo Li, Guanhua Lu and Denise Sam, thanks for

your senior leadership and being the role models for me. Konstantinos Spiliopoulos, Lucaci Vaczlavik and Cristian Tomasetti, thank you for being my long term officemates and sharing with me many great thoughts and ideas. Ritaja Sur and Anastasia Voulgaraki, thank you for friendship for the past five years, it is always great to have you there in the Stat Party. Min Tang, Lingyan Cao, Yue Tian, Wei Guo, Changhui Tan, Huashuai Qu, Minghao Wu and Neung-Soo Ha, thank you for bringing a lot of joy and cheer to my otherwise plain student life.

Many thanks to Haydee Hidalgo, Linette Berry, Celeste Regalado, Sharon Welton, Fletcher Kinne and Bill Schildknecht who put my graduate student life in order. And a loud round of applause to all the soccer fans in and out of the department who bring forward the great weakly moments to keep me in good shape for tackling the thesis.

It is impossible to remember all, and my apology to those I have inadvertently left out.

Lastly, thank you all!

CONTENTS

0. INTRODUCTION

Consider a frictionless[1] financial market in which only one risky asset (stock S_t) and one riskless asset with a constant risk free rate r are traded. To study the dynamics of the log price process $R_t = \ln S_t$, various types of models based on stochastic differential equations (SDE) have been proposed. In particular, models which incorporate stochastic volatility have entered the mainstream as scholars and market participants increasingly realize that the latent volatility is the key driving force of the market. See Fouque, Papanicolaou, and Sircar [30] and a survey study by Ghysels et al. [31] for more details. Several parametric estimation schemes have consequently been designed and tested. We refer readers to the review by Broto and Ruiz [21] on ARCH type models and the survey by Dotsis et al. [25] on SDE type models. Among the estimation techniques, likelihood based and moment based methods are the most popular choices. Although likelihood based methods are optimal when one knows the true model, they may produce biased or unstable estimators if the model specification is wrong. Also, when the marginal density of R_t does not have a closed form expression, it is impossible or computationally expensive to compute the likelihood. Moment based methods are easier to implement and

[1] A frictionless market is where all costs and restraints associated with transactions are non-existent.

less affected by model misspecification at a mild cost of efficiency. But when the number of parameters increases, the performance of moment based methods can quickly deteriorate as the higher order moments can be greatly affected by outliers and the noise in the data.

Recently, high frequency trading data have become widely available and a popular data source for parameter estimation. However, most of the research focus has been directed to estimating the variance (volatility) components of R_t (e.g. [9] and [10]), by using various types of sums of lagged (log) returns proposed by Barndorff-Nielsen and Shephard. While these efforts have resulted in many exciting advances in the study of volatility, they do not suggest how to use such data to estimate all parameters simultaneously in the model for R_t. While understanding that volatility provides deeper insight into the market, being able to characterize the dynamics of R_t is also important in different aspects of financial studies, for example, estimating risk premia (cf Broadie et al. [20]) and computing the fair value of the path dependent options.

In this paper, we try to address the above estimation problem by employing a class of well studied estimators for i.i.d. data, called Minimum Disparity Estimator (MDE). The basic idea of MDE is to minimize the distance between probabilities suggested by the model and the ones estimated from the data. The key components of the MDE are the user selected distance metric ρ, a family of parametric densities $m_\theta(x)$ indexed by θ and the kernel density estimate $f^*(x)$ computed from the data. A special class of MDE called the Minimum Hellinger Distance estimator

(MHD) has been studied by Beran [16], Tamura and Boos [71] and Simpson [65, 66]. Their results showed that MHD was robust against data contamination and model mispecification with little cost of efficency. Lindsay [49] and Basu and Lindsay [12] extended these results to general MDE for discrete and continuous models with i.i.d. data. A recent simulation study conducted by Takada [70] showed that MHD can be applied with low computation cost even when $m_\theta(x)$ has no closed form expression.

Our study focuses on the stochastic volatility model proposed by Barndorff-Nielsen and Shephard (BN-S model). We investigate one special class of MDE's called the Negative Exponential Disparity Estimates (NEDE) and apply it to estimate all of the parameters in the BN-S model simultaneously. By explicitly deriving the Taylor expansion of the Negative Exponential disparity with a special class of the BN-S model, which we have not seen in other literatures before, we obtain a concrete result on asymptotic properties of the estimator and provide the implementation details. Due to the fundamental difference between i.i.d. data and time series data and time constraint, we leave the discussion of robustness for future work.

This paper is organized as follows. In Chapter 1, we introduce the BN-S model and study how to derive the dynamics of the *Volatility Index*[2] (VIX) based on the BN-S model. We show how to facilitate the parameter estimation by using the VIX data. In Chapter 2, we prove the smoothness and differentiability of the transition and stationary density of the bivariate process (X_i, σ_i^2) derived from the BN-S model. Here, $X_i = R_i - R_{i-1}$ is the log return sequence and σ_i^2 is the squared

[2]The VIX is calculated and disseminated in real-time by the Chicago Board Options Exchange

volatility sequence. They are both observed over discrete time points. The β-mixing property of (X_i, σ_i^2) with geometric mixing rate is proved. In Chapter 3, we introduce the MDE proposed by Basu and Lindsay [12] for continuous models and study one of its special cases, called the NEDE. General results concerning the properties of the MDE are included in Appendix B.1. Technical details needed for applying the NEDE to the BN-S model are covered in Appendix B.2. Appendix B.3 discusses the functional delta method as an alternative approach to study asymptotic normality. In Chapter 4, we describe how to construct the Method of Moments (MOM) estimator had we been able to observe the latent volatility. Computations of various moments are put in Appendix C. In Chapter 5, we summarize the results and discuss some aspects for future study.

1. BN-S MODEL, EQUIVALENT MARTINGALE MEASURE (EMM) AND VIX^2 DYNAMICS

1.1 BN-S Model and the Structure Preserving EMM Transform

In this section, we formally introduce the BN-S stochastic volatility model and summarize some of the features and advantages of using this model. Then we describe the structure preserving equivalent martingale measure transform proposed by Nicolato and Venardos [55] for this model. The study of the EMM transform is of great importance to asset pricing theory, but the merit of their result to our study is that this special transform makes it straightforward to derive the dynamics of VIX^2. With the observable VIX data, we can estimate some parameters related to the volatility process.

Recall we denote the log asset value process by R_t and the squared volatility process by σ_t^2. Assume all the processes are defined on a common filtered complete probability space $(\Omega, (\mathcal{F}_t)_{0\leq t\leq T}, \mathrm{P})$ up to a finite time horizon T. Under the BN-S model, (R_t, σ_t^2) satisfies the following system of stochastic differential equations

(SDEs) under the statistical measure P:

$$
\begin{cases}
\mathrm{d}R_t \;=\; (\mu + \beta\sigma_t^2)\,\mathrm{d}t \;+\; \sigma_t\,\mathrm{d}W_t \;+\; \rho d Z_{\lambda t}, & R_0 \;=\; 0 \\
\mathrm{d}\sigma_t^2 \;=\; -\lambda\sigma_t^2\,\mathrm{d}t \;+\; \mathrm{d}Z_{\lambda t}\,, & \sigma_0^2 > 0
\end{cases}
\tag{1.1}
$$

with $\lambda > 0$ and $\rho \leq 0$, where $Z_{\lambda t}$ is the driving process with Lévy density $w(x)$ defined on $\mathbb{R}^+$ (such process is also called a subordinator) and W_t is a standard Brownian motion which is independent of Z_t. In the literature, σ_t^2 is commonly known as the (Non-Gaussian) Ornstein-Uhlenbeck (OU) process and Z_t is called the Background Driving Lévy process (BDLP).

Remark: In the original model specification, Barndorff-Nielsen and Shephard used the centered process $\bar{Z}_{\lambda t} = Z_{\lambda t} - E[Z_{\lambda t}]$ in the dynamics of R_t. Nicolato and Venardos [55] studied the Equivalent Martingale Measure for the BN-S model and they used $Z_{\lambda t}$ in the dynamics of R_t instead. As our study of the VIX^2 dynamics is based on the formulae proposed by Nicolato and Venardos, and it is clear that there is no major difference between using $Z_{\lambda t}$ or $\bar{Z}_{\lambda t}$, we will use $Z_{\lambda t}$ when specifying the model for R_t.

Remark: The BN-S model can be used to model any asset (and its volatility) traded in the market, but in order to relate the dynamic of σ_t^2 to VIX^2, we will always assume the S_t represents the S&P 500 index value.

There are several comments on the use of the OU process and the BN-S model:

- For σ_t^2:

(s1). The trajectory of σ_t^2 consists of upward jumps of Z_t with periods of downward exponential decay between them. This *asymmetric* behavior is closer to the actual behavior of volatility than the symmetric one described by constant volatility.

(s2). The mean reverting parameter λ controls the serial dependence of the process, with value close to 0 corresponding to a long memory process.

(s3). One can include more than one risk factor into σ_t^2 by superposition:

$$\sigma_t^2 = \sigma_{1t}^2 + \sigma_{2t}^2 \quad \text{where} \quad \sigma_{it}^2 = -\lambda_i \sigma_{it}^2 dt \,+\, dZ_{\lambda_i t}^i \quad \text{for } i = 1, 2.$$

Through this approach, one can include variation induced by a short-term force, such as breaking news together with influences due to long-term economic change.

(s4). The tail behavior of σ_t^2 is completely determined by the distribution of Z_1. Thus one can easily create a volatility process with heavy tail.

- For R_t:

(r1). It is common practice in finance to study quantities which depend on the unobservable volatility process, in particular the integrated volatility, through the quadratic variation of the price process. The specification of R_t in the BN-S model gives a direct connection between the quadratic variation of R_t and the integrated volatility $\int_0^h \sigma_s^2 \,\mathrm{d}s$. We will discuss this feature later in Chapter 4 to see how it helps to derive estimators for the parameters using high frequency sampled returns.

(r2). The model captures volatility feedback by $\beta\sigma_t^2$ in the expected return. For a risk-averse (or risk-neutral) market participant, β is nonnegative, meaning that the investor expects higher return with higher volatility (risk).

(r3). The model also incorporates the *leverage effect* by including the term $\rho dZ_{\lambda t}$, through which the upward jumps of σ_t^2 induce downward jumps in R_t. The strength of leverage is measured by $|\rho|$.

Next we introduce the Equivalent Martingale Measure (EMM) transform. The EMM, risk neutral measure or Q-measure, is a probability measure under which the current value of all financial assets is equal to the expected value of its future payoff when discounted by the risk-free rate. In formal mathematical language, this means that the discounted asset value $e^{-rt}S_t$ is a martingale under Q, i.e.

$$e^{-rt}S_t = E^{\mathrm{Q}}[e^{-rT}S_T|\mathcal{F}_t], \qquad \text{for} \quad T \geq t$$

where $\mathcal{F}_t$ is the σ-algebra generated by $\{S_u, u \leq t\}$. The EMM is of great importance to financial asset pricing theory, as the existence of EMM is equivalent to no arbitrage in the market (see Section 9.1 in [23] for more discussion). Therefore, when a model for a financial asset is specified, one must prove the existence of the EMM before any further analysis.

It turns out the market described by the BN-S model (1.1) is incomplete, which can be intuitively understood as saying that one has no information about the risk factor σ_t^2. In an incomplete market, the EMM is non-unique, and there are potentially infinite many EMMs (of possibly finitely many classes) for a specified model. Then the expected future payoff of the asset always equals to the risk free rate under any of these EMMs. Hubalek and Sgarra [37] studied a family of EMM transforms for the BN-S model called the Esscher transform, and they gave two approaches to characterize the change of measure. The *structure preserving equivalent martingale transform* proposed earlier by Nicolato and Venardos [55] is a special case of this family of the Esscher transform. It is called structure preserving because the independence between W_t and Z_t is preserved after the measure is changed from P (statistical) to Q (risk-neutral). Such a result is generally not true for the Esscher transform.

The structure preserving transform is of particular interest because the independence between W_t and Z_t under Q makes it possible to derive the dynamics of VIX^2 straight from its definition (cf [22] and [48]). Using the VIX data listed on CBOE, one can estimate λ and study the autocorrelation of the volatility time series. This fact is very helpful when we study the MDE and MOM estimators later. There are other advantages of this transform. For example, one can directly compare the difference between parameters before and after the change of measure, which facilitates the study of risk premia. Further, the characteristic function of R_t under Q can be easily derived and one the Fast Fourier Transform (FFT) can be used to study option pricing directly. Next we briefly summarize the result on the structure preserving equivalent martingale transform.

To present the EMM result, we need to introduce some definitions:

- Assume the filtered complete probability space $(\Omega, (\mathcal{F}_t)_{0\le t\le T}, \mathrm{P})$ satisfies the following usual hypotheses:

  (i) $\mathcal{F}_0$ contains all the P-null sets of $\mathcal{F}$;

  (ii) $\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u$, all t, $0 \le t \le T$; that is, the filtration $(\mathcal{F}_t)_{0\le t\le T}$ is right continuous.

- A stochastic process R is said to be **adapted** if $R_t \in \mathcal{F}_t$ for each t. A stochastic process R is said to be **càdlàg** if it almost surely has sample paths which are right continuous (càd), with left limits (làg).

- A process which is measurable with respect to the σ-algebra $\mathfrak{S}$ on $\Omega \times \mathbb{R}_+$ generated by all left-continuous adapted processes is called **predictable**.

- Stochastic integral with respect to Brownian motion. For a predictable càdlàg process R and Brownian motion W, $(R \cdot W)_t$ is defined as:

$$(R \cdot W)_t \triangleq \int_0^t R_{s-} d W_s = \lim_{||\pi||\to 0} \sum_{k=1}^{n} R_{t_k}\left(W_{t_{k+1}\wedge t} - W_{t_k \wedge t}\right)$$

  The limit, if exists, is understood as convergence in $L_2(P)$.

- Stochastic integral with respect to Poisson random measures μ. To simplify the discussion, we focus on the random measures for Lévy processes. Suppose a one dimensional Lévy process Z_t has discontinuity at time $T_n(\omega)$ of size

$Y_n = Z_{T_n} - Z_{T_n^-}$ for $n \geq 1$. Then its jump measure (i.e., Poisson random measure) μ_Z is defined as:

$$\mu_Z(\omega, \cdot) = \sum_{n \geq 1} \delta_{(T_n(\omega), Y_n(\omega))} = \sum_{t \in [0,T],\, \Delta Z_t \neq 0} \delta_{(t, \Delta Z_t)}.$$

Intuitively speaking, for any measurable subset of $A \subset \mathbb{R}$:

$$\mu_Z([0,t], A) := \text{number of jumps of } Z \text{ occurring between}$$
$$0 \text{ and } t \text{ whose sizes belong to } A.$$

Its compensator $\nu_Z(\cdot, \cdot)$ is given by $\nu_Z(dt, \mathrm{d}x) = dt w(\mathrm{d}x)$ where $w(\cdot)$ is the Lévy measure of Z_t. For a predictable random function $f : \Omega \times [0,T] \times \mathbb{R}^d \to \mathbb{R}$, the stochastic integral of f with respect to the compensated jump measure $(\mu_Z - \nu_Z)$ is defined as

$$\begin{aligned} f \star (\mu_z - \nu_Z) &\triangleq \int_0^T \int_{\mathbb{R}^d} f(s,y)\big(\mu_Z(ds,dy) - \nu_Z(ds,dy)\big) \\ &= \int_0^T \int_{\mathbb{R}^d} f(s,y)\big(\mu_Z(ds,dy) - w(dy)ds\big). \end{aligned}$$

Jacod [40] showed that $f \star (\mu_z - \nu_Z)$ was a martingale with respect to the time parameter t in place of T.

- $\mathcal{E}(R)$ denotes the **Stochastic Exponential** of a càdlàg process R. For a semi-martingale R, $\mathcal{E}(R)$ is defined as

$$\mathcal{E}(X) = \exp\left\{R_t - \frac{1}{2}\Big[R,R\Big]_t\right\} \prod_{0<s\leq t} (1 + \Delta R_s) \exp\left\{-\Delta R_s + \frac{1}{2}(\Delta R_s)^2\right\}$$

where $\Delta R_s = R_s - R_{s-}$ and $[R, R]_t$ is the quadratic variation process of R given by

$$[R, R]_t = \lim_{||\pi|| \to 0} \sum_{k=1}^{n} \left(R_{t_k} - R_{t_{k-1}}\right)^2$$

where π is a partition of the interval $[0, t]$ and $||\pi||$ is the mesh size. The limit, if it exists, is understood as convergence in probability.

Remark More details about these notions can be found in [23], [41] and [61].

Now we are ready to state the result by Nicolato and Venardos. Define the Cumulant Transform Function (CTF) $\kappa(\theta)$ for Z_1 as:

$$\kappa(\theta) = \log E(e^{\theta Z_1}) = \int_{\mathbb{R}^+} (e^{\theta x} - 1) w(x) \, \mathrm{d}x \tag{1.2}$$

for $\theta < \hat{\theta}$ where $\hat{\theta} = \sup\{\theta \in \mathbb{R} : \kappa(\theta) < +\infty\}$. Note that θ can be a complex number, in which case we require that $\mathbf{Re}(\theta) < \hat{\theta}$. For the given Lévy density $w(x)$, introduce a family of functions $\mathcal{Y}$:

$$\mathcal{Y} := \{y : \mathbb{R}^+ \to \mathbb{R}^+ | \int_{\mathbb{R}^+} (\sqrt{y(x)} - 1)^2 w(x) \, \mathrm{d}x < \infty\}.$$

Set $w^y(x) = y(x) w(x)$ for $y \in \mathcal{Y}$.

Lemma 1.1.1: **(Nicolato and Venardos 2003, Theorem 3.2).** Let $y \in \mathcal{Y}$. Then the process

$$\psi_t = \frac{r - \mu - (\beta + \frac{1}{2})\sigma_t^2 - \lambda \kappa^y(\rho)}{\sigma_t},$$

where $\kappa^y(\theta) = \int_{\mathbb{R}+}(e^{\theta x} - 1)w^y(x)\,\mathrm{d}x$ for $\mathbf{Re}(\theta) < 0$, is such that

$$P\left(\int_0^T \psi_s^2 ds < \infty\right) = 1$$

and

$$L_t^y = \mathcal{E}(\psi \cdot W + (y-1) \star (\mu_z - \nu_z))_t \qquad 0 \leq t \leq T$$

is a density process. The probability measure Q^y defined by $d\mathrm{Q}^y = L_T^y dP$ is an EMM and the dynamics of (R_t, σ_t^2) under Q^y are given by:

$$\begin{cases} dR_t \; = \; (r - \lambda\kappa^y(\rho) - \frac{1}{2}\sigma_t^2)dt \; + \; \sigma_t dW_t^y \; + \rho dZ_{\lambda t}^y, & \\ d\sigma_t^2 \; = \; -\lambda\sigma_t^2 dt \; + \; dZ_{\lambda t}^y, & \sigma_0^2 > 0 \end{cases} \tag{1.3}$$

where $W_t^y = W_t - \int_0^t \psi ds$ is a Q^y standard Brownian Motion and $Z_{\lambda t}^y$ is a Q^y Lévy process with Lévy density $w^y(x)$. Further, W_t^y and $Z_{\lambda t}^y$ are independent under Q^y.

Remark This lemma along with the derivation discussed in the next section will be used to find the dynamics of VIX^2 implied by the BN-S model. We focus on the cases where the BDLP Z_t is specified by the Gamma process or the Inverse Gaussian process.

Remark Another important aspect in the EMM study is the price range spanned by the value of a claim when a class of EMMs is used. We won't discuss this topic here as it is less relevant to the estimation problem in measure P. Interested readers are advised to study Chapter 5 of [55].

1.2 Deriving the Dynamics of VIX^2 implied by the BN-S Model

The key motivation to study the VIX is as follows. The purpose of VIX is to measure the market expectation of near-term future volatility conveyed by S&P 500 stock index option prices, it is natural to treat it as a proxy to study the behavior of the latent process σ_t^2. Further, notice that the mean reverting parameter λ is unchanged in the EMM transform, the result in this section shows the VIX_t^2 process also has the OU structure with exactly the same mean reverting parameter provided that the dynamics of S&P 500 index is correctly specified by the BN-S model. This suggests we can estimate λ by using the sample autocorrelation function of VIX^2. Besides, the dynamics of VIX^2 can be very useful in studying the fair value of financial derivatives which use VIX as the underlying asset, but we will not pursue this direction in our study.

Let $\mathcal{F}_t = \sigma\{(R_s, \sigma_s^2), 0 < s \le t\} \bigcup \mathcal{F}_0$. Recall the following model-free formula (definition) used by CBOE [22] to derive the current value of VIX square:

$$\text{VIX}_t^2 \triangleq \frac{2}{\tau}\sum_i \frac{\Delta K_i}{K_i^2}\tilde{V}_i(K_i) - \frac{1}{\tau}\Big[\frac{F_t(t+\tau)}{K_0} - 1\Big]^2$$

where $\tau = 30/365$, $\tilde{V}_i$ is the fair value of the out-of-the money SPX option with strike K_i, and K_0 is the highest strike below the index forward price $F_t(t+\tau)$. Lin [48] shows that this definition is a discrete approximation to

$$\frac{2}{\tau}\Big[\int_0^F \frac{\mathrm{d}K}{K^2}\tilde{P}(K) + \int_F^\infty \frac{\mathrm{d}K}{K^2}\tilde{C}(K)\Big] = -\frac{2}{\tau}E^{\mathrm{Q}^y}\Big[\ln\frac{S_{t+\tau}}{S_t e^{r\tau}} \,|\, \mathcal{F}_t\Big] \qquad (1.4)$$

where $\tilde{C}$ and $\tilde{P}$ are the forward call and put prices. Therefore, under the BN-S model, the dynamics of VIX_t^2 can be derived using the right-hand side of the above equation.

$$\begin{aligned}
\mathrm{VIX}_t^2 &= -\frac{2}{\tau} E^{\mathrm{Q}^y}\left[\ln \frac{S_{t+\tau}}{S_t e^{r\tau}} \,|\, \mathcal{F}_t\right] \\
&= -\frac{2}{\tau} E^{\mathrm{Q}^y}\left[\ln S_{t+\tau} - \ln S_t - r\tau \,|\, \mathcal{F}_t\right] \\
&= 2r - \frac{2}{\tau} E^{\mathrm{Q}^y}\left[\ln S_{t+\tau} - \ln S_t \,|\, \mathcal{F}_t\right]
\end{aligned}$$

Using the dynamical equation (1.3) of R_t (i.e., $\ln S_t$) under Q^y,

$$\begin{aligned}
&\ln S_{t+\tau} - \ln S_t \\
&\qquad = \int_t^{t+\tau}(r - \lambda\kappa^y(\rho) - \tfrac{1}{2}\sigma_s^2)ds + \int_t^{t+\tau} \sigma_s dW_s^y + \int_t^{t+\tau} \rho dZ_{\lambda s} \\
&\qquad = r\tau - \tau\lambda\kappa^y(\rho) - \tfrac{1}{2}\int_t^{t+\tau} \sigma_s^2 ds + \int_t^{t+\tau} \sigma_s dW_s^y + \rho[Z^y_{\lambda(t+\tau)} - Z^y_{\lambda t}],
\end{aligned}$$

which implies that

$$\begin{aligned}
&E^{\mathrm{Q}^y}\left[\ln S_{t+\tau} - \ln S_t | \mathcal{F}_t\right] \\
&\qquad = r\tau - \tau\lambda\kappa^y(\rho) - \frac{1}{2} E^{\mathrm{Q}^y}\left[\int_t^{t+\tau} \sigma_s^2 ds | \mathcal{F}_t\right] \\
&\qquad\qquad + E^{\mathrm{Q}^y}\left[\int_t^{t+\tau} \sigma_s dW_s^y | \mathcal{F}_t\right] + \rho E^{\mathrm{Q}^y}\left[Z^y_{\lambda(t+\tau)} - Z^y_{\lambda t} | \mathcal{F}_t\right] \\
&\qquad = r\tau - \tau\lambda\kappa^y(\rho) - \frac{1}{2} E^{\mathrm{Q}^y}\left[\int_t^{t+\tau} \sigma_s^2 ds | \mathcal{F}_t\right] + \lambda\,\rho\,\tau E^{\mathrm{Q}^y}[Z_1^y]
\end{aligned}$$

The last equality is due to the time homogeneous property of a Lévy process. Finally

$$\begin{aligned}
&-\frac{2}{\tau} E^{\mathrm{Q}^y}\left[\ln S_{t+\tau} - \ln S_t | \mathcal{F}_t\right] \\
&\qquad = -2r + 2\lambda\kappa^y(\rho) + \frac{1}{\tau} E^{\mathrm{Q}^y}\left[\int_t^{t+\tau} \sigma_s^2 ds | \mathcal{F}_t\right] - 2\lambda\rho E^{\mathrm{Q}^y}[Z_1]
\end{aligned}$$

If we assume that the Lévy process Z_t has finite mean under both P and Q^y, then

$$E^{\mathrm{Q}^y}[Z_1] \quad = \quad \tfrac{\partial}{\partial\theta}\kappa^y(\theta)\Big|_{\theta=0} \quad = \quad \int_{\mathbb{R}+} xy(x)w(x)\,\mathrm{d}x\,.$$

Further, for $s > t$, by using integration by parts, one can derive

$$\sigma_s^2 = e^{-\lambda(s-t)}\sigma_t^2 + \int_t^s e^{-\lambda(s-u)}\,\mathrm{d}Z_{\lambda u}^y$$

and then

$$\begin{aligned}
&E^{\mathrm{Q}^y}[\int_t^{t+\tau} \sigma_s^2 ds|\mathcal{F}_t]\\
&= E^{\mathrm{Q}^y}[\int_t^{t+\tau} e^{-\lambda(s-t)}\,\mathrm{d}s\ \sigma_t^2 + \int_t^{t+\tau}\int_t^s e^{-\lambda(s-u)}\,\mathrm{d}Z_{\lambda u}^y\,\mathrm{d}s\,|\mathcal{F}_t]\\
&= \tfrac{1}{\lambda}[e^{-\lambda(t-t)} - e^{-\lambda(t+\tau-t)}]\cdot\sigma_t^2 + E^{\mathrm{Q}^y}[Z_1]\cdot\int_t^{t+\tau}\int_t^s e^{-\lambda(s-u)}\lambda\ \mathrm{d}u\,\mathrm{d}s\\
&= \tfrac{1}{\lambda}[1-e^{-\lambda\tau}]\cdot\sigma_t^2 + E^{\mathrm{Q}^y}[Z_1]\cdot[\tau - \tfrac{1}{\lambda}(1-e^{-\lambda\tau})]
\end{aligned}$$

Thus, the $-\frac{2}{\tau}$ normalized conditional expectation of the log return under Q^y is given by:

$$\begin{aligned}
-\frac{2}{\tau}E^{\mathrm{Q}^y}\left[\ln S_{t+\tau} - \ln S_t|\mathcal{F}_t\right] = -2r + \frac{1-e^{-\lambda\tau}}{\lambda\tau}\sigma_t^2 + \left(1 - \frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)E^{\mathrm{Q}^y}[Z_1]\\
+ 2\lambda\int_{\mathbb{R}+}(e^{\rho x} - 1 - \rho x)y(x)w(x)\,\mathrm{d}x\,.
\end{aligned}$$

Therefore, under the measure Q^y, VIX_t^2 is given by:

$$\mathrm{VIX}_t^2 = \frac{1-e^{-\lambda\tau}}{\lambda\tau}\sigma_t^2 + D(\lambda,\tau,y(\cdot),w(\cdot)) \tag{1.5}$$

with $D(\lambda,\tau,y(\cdot),w(\cdot)) = 2\lambda\int_{\mathbb{R}+}(e^{\rho x}-1-\rho x)y(x)w(x)\,\mathrm{d}x + (1-(1-e^{-\lambda\tau})/(\lambda\tau))E^{\mathrm{Q}^y}[Z_1]$. If one chooses $\sigma_0^2 \overset{\mathfrak{D}}{=} \int_0^\infty e^{-t}\,\mathrm{d}Z_t$ as we do from this point on, then σ_t^2 is strictly stationary[1], which implies VIX_t^2 is also strictly stationary. Let

$$V_t \triangleq \mathrm{VIX}_t^2 - D(\lambda, \tau, y(\cdot), w(\cdot)).$$

Then V_t satisfies the following SDE under Q^y:

$$dV_t = -\lambda V_t dt + \tfrac{1-e^{-\lambda\tau}}{\lambda\tau} dZ^y_{\lambda t}, \qquad V_0 \stackrel{\mathcal{D}}{=} \tfrac{1-e^{-\lambda\tau}}{\lambda\tau} \int_0^\infty e^{-t} dZ^y_t .$$

One can also derive the characteristic function for VIX_t^2:

$$\begin{aligned} \phi_{\mathrm{VIX}_t^2}(u) = E^{\mathrm{Q}^y}\left[e^{iu(V_t+D)}\right] &= e^{iuD} \cdot E^{\mathrm{Q}^y}\left[e^{iu\frac{1-e^{-\lambda\tau}}{\lambda\tau}\sigma_t^2}\right] \\ &= e^{iuD} \cdot \phi^y_{\sigma_t^2}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} u\right) \\ &= e^{iuD} \cdot \phi^y_{\sigma_0^2}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} u\right) \end{aligned}$$

where $\phi^y_{\sigma_t^2}(u)$ is the characteristic function of σ_t^2 under Q^y.

From (1.5), one immediately gets the following properties for the moments of VIX_t^2:

- $E^{\mathrm{Q}^y}[\mathrm{VIX}_t^2] = \frac{1-e^{-\lambda\tau}}{\lambda\tau} E^{\mathrm{Q}^y}[\sigma_t^2] + D$;
- $\mathrm{Var}^{\mathrm{Q}^y}[\mathrm{VIX}_t^2] = (\frac{1-e^{-\lambda\tau}}{\lambda\tau})^2 \mathrm{Var}^{\mathrm{Q}^y}[\sigma_t^2]$;
- $\rho^{\mathrm{Q}^y}[\mathrm{VIX}_t^2, \mathrm{VIX}_s^2] = \rho^{\mathrm{P}}[\sigma_t^2, \sigma_s^2] = e^{-\lambda|t-s|}$.

From the last equation, one finds that if the VIX index accurately approximates the left-hand side of (1.4), then the mean reverting parameter λ can be estimated by using the VIX^2 data.

[1]See Lemma 2.1.1 and the following discussion.

1.3 Examples of Structure Preserving EMM for the BN-S Model

In this section, we will study some analytic properties of two OU processes: the Gamma OU (Γ-OU) process and the Inverse Gaussian OU (IG-OU) process. We choose these two processes because, along with the Tempered Stable process, they are the most analytically tractable pure jump processes with only positive jumps. Besides, the Gamma OU process can be simulated very efficiently and is therefore a good candidate for a simulation study. Further, empirical studies (cf [7]) have shown that the distribution of volatility can be well approximated by the Inverse Gaussian distribution.

We focus on the following three aspects: the characteristic functions of the stationary distributions of these two processes, the corresponding structure preserving EMM transform and the VIX_t^2 dynamics. First we review some basics of the Lévy-Khintchine formula (cf [63]).

For any Lévy process Z_t, the distribution F of Z_1 is infinitely divisible. The **Lévy - Khintchine decomposition formula** states that the characteristic function of any infinitely divisible distribution can always be written in the following form (when Z_1 is univariate):

$$\phi_{Z_1}(u) = \exp\left[i\gamma u - \frac{\sigma^2}{2}u^2 + \int_{\mathbb{R}}(e^{iux} - 1 - iux\mathbb{I}_{\{|x|<1\}})\Pi(\mathrm{d}x)\right],$$

where $\gamma \in \mathbb{R}$, $\sigma^2 \geq 0$ and $\Pi(\cdot)$ is a measure on $\mathbb{R}$ with

$$\int_{\mathbb{R}} (1 \wedge x^2)\Pi(\,\mathrm{d}x\,) < \infty.$$

We say that $[\gamma, \sigma^2, \Pi(\,\mathrm{d}x\,)]$ is the Lévy triplet of Z_1 and $\Pi(\cdot)$ is called the Lévy measure of Z_t. If $\Pi(\cdot)$ has a density $w(x)$ with respect to the Lebesgue measure, we also refer to $[\gamma, \sigma^2, w(x)]$ as the Lévy triplet.

Since in the BN-S model, Z_t is a subordinator and has positive jumps only, then $w(\cdot)$ is defined only on $\mathbb{R}^+$ and $\int_0^1 xw(x)\,\mathrm{d}x\, < \infty$ because Z_t has finite variation. The characteristic function Z_1 is simplified to

$$\phi_{Z_1}(u) = \exp\left\{\int_{\mathbb{R}+}(e^{iux} - 1)w(x)\,\mathrm{d}x\right\} \quad \text{for} \quad u \in \mathbb{R}.$$

(1). BN-S model with $\Gamma(\nu, \alpha)$-OU Volatility Process:

For a compound Poisson process Z_t with Lévy density $w(x)$, we know its Lévy triplet is given by:

$$\left[\int_{-1}^{1} xw(x)\,\mathrm{d}x\,,\, 0\,,\, w(x)\right].$$

In the **Γ**-OU case, the BDLP Z_t is a Compound Poisson process with Lévy density $w(x) = \nu\alpha e^{-\alpha x}$ for $x > 0$. So it is easy to obtain that Z_t has Lévy triplet

$$\left[\frac{(1-e^{-\alpha}(1+\alpha))\nu}{\alpha}, 0, \nu\alpha e^{-\alpha x}\right],$$

and its CTF is $\kappa(\theta) = \frac{\nu\theta}{\alpha-\theta}$. It can be shown that σ_t^2 is a stationary process whose marginal distribution is Gamma(ν, α). Thus σ_t^2 is also a Lévy process with the following Lévy triplet:

$$\left[\frac{\nu}{\alpha}(1-e^{-\alpha}),\ 0\,,\ \frac{\nu}{x}e^{-\alpha x}\mathbb{I}_{\{x>0\}}\right],$$

and its CTF is given by $\kappa^{\Gamma}(\theta) = \ln\left[\left(\frac{\alpha}{\alpha-\theta}\right)^{\nu}\right]$ for $\mathbf{Re}(\theta) < \alpha$. Define the following processes:

$$\begin{cases} y(x) &=& \frac{\tilde{\nu}\tilde{\alpha}}{\nu\alpha}e^{-(\tilde{\alpha}-\alpha)x}, \qquad \text{for } \tilde{\nu},\tilde{\alpha}\in\mathbb{R}^{+} \\ \kappa^{y}(\theta) &=& \frac{\tilde{\nu}\rho}{\tilde{\alpha}-\rho} \\ \psi_t &=& \frac{r-\mu-(\beta+1/2)\sigma_t^2-\lambda(\tilde{\nu}\rho)/(\tilde{\alpha}-\rho)}{\sigma_t}. \end{cases}$$

Let μ_Z denote the jump measure of Z_t and $\nu_Z(x,t)$ denote its compensator (in this case, $d\nu_Z(x,t) = \lambda\nu\alpha e^{-\alpha x}\,\mathrm{d}x\,dt$). Then, according to Corollary (3.3) in [55], the process $L_t^y = \mathcal{E}[\psi\cdot W_t + (y(x)-1)\star(\mu_Z-\nu_Z)]_t$ $0\le t\le T$ is a density. The EMM transform which preserves the BN-S structure is given by $d\mathrm{Q}^y = L_T^y dP$. By (1.5),

$$\mathrm{VIX}_t^2 = \frac{1-e^{-\lambda\tau}}{\lambda\tau}\sigma_t^2 + \left[2\lambda\frac{\tilde{\nu}\rho^2}{\tilde{\alpha}^2-\tilde{\alpha}\rho} + \left(1-\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)\cdot\frac{\tilde{\nu}}{\tilde{\alpha}}\right].$$

Using (C.4), one can compute the following three moments (cross-moment):

- $E^{\mathrm{Q}^y}[\mathrm{VIX}_t^2] \;=\; 2\lambda\frac{\tilde{\nu}\rho^2}{\alpha^2-\alpha\rho} + \frac{\tilde{\nu}}{\tilde{\alpha}}$;
- $\mathrm{Var}^{\mathrm{Q}^y}[\mathrm{VIX}_t^2] \;=\; \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)^2\cdot\frac{\tilde{\nu}}{\tilde{\alpha}^2}$;
- $\mathrm{Cov}^{\mathrm{P}}[R_{t+h}-R_t, \sigma_{t+h}^2-\sigma_t^2] \;=\; \rho(1-e^{-\lambda h})\frac{2\nu}{\alpha^2}$.

(2). BN-S Model with $IG(\delta,\gamma)$-OU Volatility Process:

Selected properties of the Inverse Gaussian (IG) distribution (following the notation in [55]):

(1) The density of the IG(δ, γ) distribution is given by:

$$f(x) = \frac{\delta}{\sqrt{2\pi}} e^{\delta\gamma} x^{-3/2} \exp\left[-\frac{1}{2}(\delta^2 x^{-1} + \gamma^2 x)\right] \quad \text{for} \quad x > 0.$$

The parameters δ and γ are positive.

(2) If the random variable X_1 follows an IG(δ, γ) distribution, then by the Lévy -Khintchine formula, its characteristic function is given by:

$$\begin{aligned} \phi_{X_1}(u) &= \exp\left(\delta(\gamma - \sqrt{\gamma^2 - 2iu})\right) \\ &= \exp\left(\delta\gamma - \frac{\delta}{\sqrt{2}}\sqrt{\sqrt{\gamma^4 + 4u^2} + \gamma^2} + i\frac{\delta}{\sqrt{2}}\sqrt{\sqrt{\gamma^4 + 4u^2} - \gamma^2}\right) \end{aligned}$$

The last equality follows from the square root formula for complex numbers. One can also derive its Lévy triplet:

$$\left[\frac{\delta}{\gamma}(2\Phi(\gamma) - 1)\,,\, 0\,,\, \frac{1}{\sqrt{2\pi}}\delta x^{-3/2} \exp\left(-\frac{\gamma^2 x}{2}\right)\mathbb{I}_{x>0}\right],$$

where $\Phi(\cdot)$ is the cumulative distribution function for a standard normal random variable.

(3) An IG(δ, γ) random variable has CTF $\kappa^{IG}(\theta) = \delta\gamma - \delta(\gamma^2 - 2\theta)^{1/2}$ and MGF $M(\theta) = e^{\delta\gamma - \delta(\gamma^2 - 2\theta)^{1/2}}$ defined for all $\theta \in (-\infty, \gamma^2/2)$.

(4) The IG(δ, γ) distribution is self-decomposable[2].

Some basic properties of the IG(δ, γ)-OU process:

[2]A random variable (or equivalently, its distribution) is self-decomposable if its characteristic function $\phi(u)$ satisfies $\phi(u) = \phi(c^{-1}u)\phi_c(u)$ for some $c > 1$ and $\phi_c(u)$ is the characteristic function for some distribution.

(1) If the BDLP Z_t in the BN-S model has the following Lévy density:

$$w(x) = \frac{\delta}{2\sqrt{2\pi}} x^{-3/2}(1+\gamma^2 x)e^{-\frac{1}{2}\gamma^2 x} \quad \text{for} \quad x > 0,$$

with $\delta > 0$ and $\gamma \geq 0$, then σ_t^2 is a stationary OU process with IG(δ, γ) marginal distribution;

(2) Z_1 has CTF $\kappa(\theta) = \theta\delta(\gamma^2 - 2\theta)^{-\frac{1}{2}}$, which is well defined for $\mathbf{Re}(\theta) < \gamma^2/2$;

(3) In the BN-S model with volatility assumed to be an IG(δ, γ)-OU process, if $\rho = 0$, then the log return R_t is approximately Normal Inverse Gaussian (NIG) distributed. The NIG distribution has density function

$$g(x;\alpha,\beta,\mu,\delta) = a(\alpha,\beta,\mu,\delta) q\left(\frac{x-\mu}{\delta}\right)^{-1} K_1\left\{\delta\alpha q\left(\frac{x-\mu}{\delta}\right)\right\} \exp(\beta x)$$

where $a(\alpha,\beta,\mu,\delta) = \alpha/\pi \, \exp(\delta\sqrt{\alpha^2-\beta^2} - \beta\mu)$, $q(x) = \sqrt{1+x^2}$ and K_1 is the modified Bessel function of the third kind wiht index 1. Furthermore, α, β, μ and δ satisfy $0 \leq |\beta| \leq \alpha$, $\mu \in \mathbb{R}$ and $0 < \delta$. Barndorff-Nielsen [4] studied how the NIG distribution captured the important empirical phenomena of stock return data.

According to Corollary 3.3 in [55], the set M^{IG} of EMMs which preserves the IG-OU structure is given by:

$$M^{IG} = \{\mathbb{Q}^y \in M' : \; y(x) = \tfrac{1+\tilde{\gamma}^2 x}{1+\gamma x} \exp\Big[-\tfrac{1}{2}(\tilde{\gamma}^2 - \gamma^2)x\Big], \quad \text{for } \tilde{\gamma} \in \mathbb{R}^+\}$$

Here, M' is the set of EMMs where the structure of the SDEs (1.1) is preserved after the transform (possibly with different parameters). With the $y(x)$ specified in $\boldsymbol{M}^{IG}$, one can find the ψ_t in the same way as in the Γ-OU case, which leads to the appropriate density process. Notice that in order to preserve the BN-S structure, the coefficient δ is the same under P and Q^y. Under Q^y, Z_1 has CGF $\kappa^y(\theta) = \theta\delta(\tilde{\gamma}^2 - 2\theta)^{-1/2}$ and σ_0^2 has CGF $\kappa_D(\theta) = \delta\tilde{\gamma} - \delta(\tilde{\gamma}^2 - 2\theta)^{1/2}$. By using the Lévy density of Z_1 under Q^y, we can compute the following quantities used in formula (1.5):

- $E^{\mathrm{Q}^y}[Z_1] = \delta\tilde{\gamma}^{-1}$;
- $\mathrm{Var}^{\mathrm{Q}^y}[Z_1] = 2\delta\tilde{\gamma}^{-3}$;
- $\int_{\mathbb{R}+}(e^{\rho x} - 1 - \rho x)y(x)w(x)\ \mathrm{d}x = \rho\delta\left(\frac{1}{\sqrt{\tilde{\gamma}^2 - 2\rho}} - \frac{1}{\tilde{\gamma}}\right)$.

Thus VIX_t^2 under Q^y can be expressed as:

$$\mathrm{VIX}_t^2 = \frac{1-e^{-\lambda\tau}}{\lambda\tau}\sigma_t^2 + 2\lambda\rho\delta\left(\frac{1}{\sqrt{\tilde{\gamma}^2-2\rho}} - \frac{1}{\tilde{\gamma}}\right) + \left(1 - \frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)\cdot\frac{\delta}{\tilde{\gamma}}.$$

Since $\sigma_t^2 \sim \mathrm{IG}(\delta, \tilde{\gamma})$, $E^{\mathrm{Q}^y}[\sigma_t^2] = \delta/\tilde{\gamma}$ and $\mathrm{Var}^{\mathrm{Q}^y}[\sigma_t^2] = \delta/\tilde{\gamma}^3$, we have the moments of VIX_t^2:

- $E^{\mathrm{Q}^y}[\mathrm{VIX}_t^2] = 2\lambda\rho\delta\left(\frac{1}{\sqrt{\tilde{\gamma}^2-2\rho}} - \frac{1}{\tilde{\gamma}}\right) + \frac{\delta}{\tilde{\gamma}}$;
- $\mathrm{Var}^{\mathrm{Q}^y}[\mathrm{VIX}_t^2] = \left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right)^2\cdot\frac{\delta}{\tilde{\gamma}^3}$;
- $\mathrm{Cov}^{\mathrm{P}}[R_{t+h} - R_t, \sigma_{t+h}^2 - \sigma_t^2] = \rho(1 - e^{-\lambda h})\cdot\frac{2\delta}{\gamma^3}$.

2. SMOOTHNESS OF TRANSITION DENSITY, MARGINAL DENSITY AND ERGODICITY

In the previous section we have modeled the VIX_t^2 process as an affine transform of σ_t^2 with the similar OU structure. Since the moments of VIX_t^2 (σ_t^2) and R_t are relatively easy to compute, it is natural to estimate the parameters in the BN-S model by the method of moments. Although we can observe VIX_t^2 and R_t at discrete time points T_i, we cannot observe the latent process σ_t^2. This suggests methods based only on the sampled return $X_i = R_{T_i} - R_{T_{i-1}}$ are needed to estimate all the parameters in BN-S model under the statistical measure P. Notice that the time series $\{X_i\}$ is a sequence of dependent variables, so that extra conditions need to be imposed on the covariance for making statistical inference. One of the common assumptions is that the series has a strong mixing property (see [26] for general discussion on various types of mixing notions). In this section, we will prove $\{(X_i, \sigma_i^2)\}$ is β-mixing with geometric mixing rate (thus it is strong mixing). As an application, we will use this conclusion to show the consistency and asymptotic normality of the MDE and MOM estimator. Also, we prove the smoothness of the density of X_i. This property is useful for computing the kernel density estimate.

Assume that we observe $(N+1)$ pairs of data (R_i, σ_i^2) from (R_t, σ_t^2) on equi-spaced time points $T_i = iT/N$ for $i = 0, 1, \dots, N$. Let X_i be the discrete time increment process given by $X_i = R_{T_i} - R_{T_{i-1}}$ and σ_i^2 is the squared spot-volatility process defined as $\sigma_i^2 = \sigma_{T_i}^2$. The joint dynamics of (X_i, σ_i^2) under P can be described by the following system of equations:

$$\begin{cases} X_i \;=\; \mu h + \beta \int\limits_{(i-1)h}^{ih} \sigma_s^2 \, \mathrm{d}s \;+\; \int\limits_{(i-1)h}^{ih} \sigma_s \, \mathrm{d}W_s \;+\; \rho \int\limits_{(i-1)h}^{ih} \mathrm{d}Z_{\lambda s}\,, & X_0 \;= 0 \\ \sigma_i^2 \;=\; e^{-\lambda h} \sigma_{i-1}^2 \;+\; \int\limits_{(i-1)h}^{ih} e^{-\lambda(ih-s)} \, \mathrm{d}Z_{\lambda s} & \sigma_0^2 \;\stackrel{\mathcal{D}}{=} \int\limits_0^{\infty} e^{-s} \, \mathrm{d}Z_s \end{cases} \tag{2.1}$$

We choose this particular combination of the increment and spot process for the following reasons:

(1) R_{T_i} (or R_t) itself is not a stationary process, whereas the increment process X_i is stationary. Besides, the log return $\{R_{T_i} - R_{T_{i-1}}\}$ is a more commonly studied process in empirical finance;

(2) X_i alone is *NOT* a Markov chain, which excludes the use of powerful techniques based on the Markov assumption;

(3) If under the statistical measure one can establish an affine relation between σ_i^2 and other observable quantities, such as trading volume, then one can take advantage of the joint mixing property of (X_i, σ_i^2) and estimate parameters more efficiently, (cf Hubalek and Posedel [36]).

The main machinery we employ is the Foster-Lyapunov type geometric ergodicity criterion proposed by Nummelin and Tuominen [56]. In order to apply this criterion, we first need to show that the following two properties hold for (X_i, σ_i^2) under certain conditions:

1. (X_i, σ_i^2) is a (strictly) stationary Markov chain where the support of its stationary distribution F has a non-empty interior;

2. The transition semigroup $\mathcal{P}_n$ for (X_i, σ_i^2) has the weak Feller property (the definition is given in Section 2.2);

This chapter is organized as follows. First, we study the Markov property of (X_i, σ_i^2) and show that this bivariate process is strictly stationary with some stationary distribution F if a proper initial distribution is chosen. Second, we study the smoothness of the transition and stationary probability measure. As a consequence, the Strong Feller property for $\mathcal{P}_n$ is proved. At last we apply the theorem in [56] to prove that (X_i, σ_i^2) is β-mixing with geometric mixing rate.

2.1 Markov Property of (X_i, σ_i^2)

The Markov property of (X_i, σ_i^2) is readily established due to the BN-S model specification: for any bounded function $f(\cdot,\cdot)$ defined on $\mathcal{B}(\mathbb{R}, \mathbb{R}^+)$, we have:

$$
\begin{aligned}
& E(f(X_i, \sigma_i^2) | X_{i-1}, X_{i-2}, \ldots, X_1; \sigma_{i-1}^2, \sigma_{i-1}^2, \ldots, \sigma_1^2) \\
&= E(f(X_i, \sigma_i^2) | X_{i-1}, \sigma_{i-1}^2) \\
&= E(f(X_i, \sigma_i^2) | \sigma_{i-1}^2)
\end{aligned}
$$

since the behavior of X_i and σ_i^2 depend only on σ_{i-1}^2 and the trajectories of

W_s and Z_s for $s \in ((i-1)h, ih]$. To justify the last equality, notice that from (2.1):

$$X_i = \mu h + \beta\Big(\int_{(i-1)h}^{ih} e^{-s}\sigma^2_{(i-1)h}\,\mathrm{d}s + \int_{(i-1)h}^{ih} e^{-s} \int_{(i-1)h}^{s} \mathrm{d}Z_{\lambda u}\,\mathrm{d}s \Big)$$
$$+ \int_{(i-1)h}^{ih} \sigma_s\,\mathrm{d}W_s + \rho \int_{(i-1)h}^{ih} \mathrm{d}Z_{\lambda s}\,,$$

Using the fact that W_s and Z_s are processes with independent increments, one finds X_i does not depends on X_{i-1}.

To prove the strict stationarity of (X_i, σ_i^2), we will use a lemma concerning the strict stationarity of σ_t^2. First let us introduce the following terminology: for a random variable X having characteristic function $\phi_X(u)$, its *characteristic exponent* is defined as $\psi_X(u) = \ln \phi_X(u)$.

Remark Sato and Yamazato used the term *characteristic exponent* in this lemma as their work is based on the characteristic function (or Fourier transform) of the density function. Compared to the CFT defined in Section (1.1), the *Cumulant Transform Function* is based on the Laplace transform of a density function.

The following lemma, restated in our notations, provides a sufficient condition for σ_t^2 to be strictly stationary.

Lemma 2.1.1: **(Sato and Yamazato 1984, Theorem 4.1 and 4.2).** Consider the volatility process σ_t^2 in the BN-S model and let $\mathcal{S}$ and $\mathcal{B}(\mathcal{S})$ denotes its sample space and the Borel σ-algebra generated by $\mathcal{S}$ respectively. Define the transition probability $P_t(x, A) \triangleq \mathrm{P}(\sigma^2_{s+t} \in A | \sigma^2_s = x)$ with $x \in \mathcal{S}$ and $A \in \mathcal{B}(\mathcal{S})$. Let the

Lévy triplet of Z_1 be $(\gamma, 0, \Pi)$. Then the following two statements hold:

(a) Let $\lambda > 0$. If

$$\int_{x>1} \log x \, \Pi(\,\mathrm{d}x\,) < \infty, \tag{2.2}$$

then there exists a limiting distribution $F_{\sigma_0^2}$ such that

$$P_t(x, A) \to F_{\sigma_0^2}(A), \qquad \text{as } t \to \infty$$

for any $x \in \mathcal{S}$ and $A \in \mathcal{B}(\mathcal{S})$. This $F_{\sigma_0^2}$ is self-decomposable and the unique invariant distribution of σ_t^2. Moreover, the characteristic function of $F_{\sigma_0^2}$ is given by

$$\phi_{\sigma_0^2}(u) = \exp\left(\int_0^\infty \psi_{Z_1}(e^{-s}u)ds\right).$$

In particular, the Lévy triplet of σ_0^2 is given by $[\gamma_{\sigma_0^2}, 0, \Pi_{\sigma_0^2}]$, where

$$\gamma_{\sigma_0^2} = \frac{\gamma}{\lambda} + \int_{\mathbb{R}}\int_0^\infty e^{-\lambda s}x\left(\mathbb{I}_{\{|e^{-\lambda s}x|<1\}} - \mathbb{I}_{\{|z|<1\}}\right)ds\Pi(\,\mathrm{d}x\,),$$

$$\Pi_{\sigma_0^2}(E) = \int_0^\infty \Pi(e^{\lambda s}E)ds, \qquad E \in \mathcal{B}(\mathbb{R}).$$

Here ψ_{Z_1} is the characteristic exponent of Z_1 .

(b) Let $\lambda \in \mathcal{S}$. If (2.2) fails to hold, then σ_t^2 has no invariant distribution, and moreover, for any $x \in \mathcal{S}$, $P_t(A, x)$ does not converge to any probability measure as $t \to \infty$.

One sees that by assuming (2.2) and choosing $\sigma_0^2 \overset{\mathfrak{D}}{=} \int_0^\infty e^{-s}\,\mathrm{d}Z_s$, the unique invariant distribution, the continuous time process σ_t^2 as well as the discrete time process σ_i^2 are strictly stationary with marginal distribution $F_{\sigma_0^2}$. This implies the sequence of *Integrated Volatility* $\int_{(i-1)h}^{ih} \sigma_s^2 \, \mathrm{d}s$ and $\int_{(i-1)h}^{ih} \mathrm{d}Z_{\lambda s}$ on successive time intervals $[T_{i-1}, T_i]$ both form strictly stationary time series. Thus we find X_i is also a strictly stationary process. Putting these results together, $\{(X_i, \sigma_i^2), i = 1, 2 \ldots N\}$ is a strictly stationary Markov chain with stationary distribution being the joint distribution F of (X_1, σ_1^2).

2.2 *Weak Feller Property of the Transition Semigroup* $(\mathcal{P}_n)_{n\in\mathbb{Z}_+}$

Following are some symbols to be used in this section:

- $b\mathcal{B}(\mathcal{S})$: space of bounded and $\mathcal{B}(\mathcal{S})$ measurable functions.
- $C_b(\mathcal{S})$: space of functions f defined on $\mathcal{S}$ which are bounded and continuous.
- $C_c^\infty(\mathcal{S})$: space of functions f defined on $\mathcal{S}$ which are infinitely many times differentiable and have compact support.
- Essential supremum norm $\| \, \|_\infty$ on functions:

$$\| f \|_\infty := \inf\{C \geq 0 : |f(x)| \leq C \text{ for almost all } x \text{ in its support } \}.$$

For the discrete time Markov chain (X_i, σ_i^2), there is an associated transition semigroup $(\mathcal{P}_n)_{n\in\mathbb{Z}_+}$ with the 1–step transition operator $\mathcal{P}_1$ defined by:

$$\begin{aligned}\mathcal{P}_1 f(x,v) &= E[f(X_i,\sigma_i^2)|\{X_{i-1}=x,\sigma_{i-1}^2=v\}]\\ &= \textstyle\int_{\mathbb{R}}\int_{\mathbb{R}^+} f(y_1,y_2)\,P_1(\mathrm{d}\mathbf{y}\,,v)\end{aligned}$$

for any bounded $f:\mathbb{R}\times\mathbb{R}^+\to\mathbb{R}$, where $P_1(A;v)\triangleq\mathbb{P}((X_i,\sigma_i^2)\in A|\sigma_{i-1}=v)$ for $A\in\mathcal{S}$ is the 1-step transition probability measure. Recall $(\mathcal{P}_n)_{n\in\mathbb{Z}_+}$ (resp. $(\mathcal{P}_t)_{t\in\mathbb{R}_+}$) is called *Weak Feller* if $\mathcal{P}_n f\in C_b(\mathcal{S})$ (resp. $\mathcal{P}_t f\in C_b(\mathcal{S})$) for any $f\in C_b(\mathcal{S})$. To show the Weak Feller property for the semigroup $(\mathcal{P}_n)_{n\in\mathbb{Z}_+}$, it suffices to show that $\mathcal{P}$ has the Weak Feller property. In the rest of the section, we will suppress the subscript n unless stated otherwise. One sees that the value of $\mathcal{P}f(x,v)$ depends on v only, so that $\mathcal{P}f(x,v)$ is bounded and continuous in x automatically. Therefore, it is only necessary to show, for $v_1,v_2\in\mathbb{R}^+$, that

$$\mathcal{P}f(x,v_2)\to\mathcal{P}f(x,v_1)\quad\text{as } v_2\to v_1.$$

Theorem 2.2.1: Under the BN-S model, the transition operator $\mathcal{P}$ for (X_i,σ_i^2) is (weak) Feller.

Proof: For v_1, $v_2\in\mathbb{R}^+$:

$$\begin{aligned}&|\mathcal{P}f(x,v_1)-\mathcal{P}f(x,v_2)|\\ &=\left|E\Big[f(X_i,\sigma_i^2)|\{X_{i-1}=x,\sigma_{i-1}^2=v_1\}\Big]-E\Big[f(X_i,\sigma_i^2)|\{X_{i-1}=x,\sigma_{i-1}^2=v_2\}\Big]\right|\\ &=\left|E\Big[E\big[f(X_i,\sigma_i^2)|\sigma_{i-1}^2=v_1\big]\Big|\sigma\{Z_s\}_{(i-1)h<s\le ih}\Big]\right.\\ &\qquad\left.-E\Big[E\big[f(X_i,\sigma_i^2)|\sigma_{i-1}^2=v_2\big]\Big|\sigma\{Z_s\}_{(i-1)h<s\le ih}\Big]\right|\end{aligned}$$

After conditioning on $\sigma\{Z_s\}_{(i-1)h<s\leq ih}$, the random element in X_i is the stochastic integral $\int_{(i-1)h}^{ih} \sigma_s dW_s$. It is easy to see that $\int_{(i-1)h}^{ih} \sigma_s dW_s$ is a Normal random variable with mean 0 and variance $\sigma_h^{*2}(v)$:

$$\begin{aligned}\sigma_h^{*2}(v) \triangleq \int_{(i-1)h}^{ih} \sigma_s^2 \, \mathrm{d}s \;&=\; \int_{(i-1)h}^{ih} e^{-\lambda s+\lambda(i-1)h} \, \mathrm{d}s \; v + \int_{(i-1)h}^{ih} \int_{(i-1)h}^{s} e^{-\lambda(s-u)} \, \mathrm{d}Z_{\lambda u} \, \mathrm{d}s \\ &=\; \frac{1}{\lambda}(1-e^{-\lambda h})v + \frac{1}{\lambda}\int_{(i-1)h}^{ih} [1-e^{-\lambda(ih-u)}] \, \mathrm{d}Z_{\lambda u}\end{aligned}$$

Since $v_1,\, v_2 \in \mathbb{R}^+$, the variance $\sigma_h^{*2}(v)$ is always strictly positive. Further, let $\sigma_{i,v}^2 = e^{-\lambda h}v + \int_{(i-1)h}^{ih} e^{-\lambda u} \, \mathrm{d}Z_{\lambda u}$ and define function $A(z,v)$ as

$$A(z,v) \;\triangleq\; \frac{\left[z - \left(\mu h + \beta\sigma_h^{*2}(v) + \rho\int_{(i-1)h}^{ih} dZ_{\lambda u}\right)\right]^2}{2\sigma_h^{*2}(v)}.$$

Then by conditioning and expressing in terms of the normal density function,

$$\begin{aligned}&|\mathcal{P}f(x,v_1) - \mathcal{P}f(x,v_2)| \\ &= \Bigg| E\Big[\int_{\mathbb{R}} f(z,\sigma_{i,v_1}^2)\cdot\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)} \, \mathrm{d}z\Big] \\ &\quad - E\Big[\int_{\mathbb{R}} f(z,\sigma_{i,v_2}^2)\cdot\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_2)}}\cdot e^{-A(z,v_2)} \, \mathrm{d}z\Big]\Bigg| \\ &\leq E\Big[\int_{\mathbb{R}} \big|f(z,\sigma_{i,v_1}^2) - f(z,\sigma_{i,v_2}^2)\big|\cdot\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)} \, \mathrm{d}z\Big] \\ &\quad + E\Big[\int_{\mathbb{R}} \big|f(z,\sigma_{i,v_2}^2)\big|\cdot\Big|\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_2)}}\cdot e^{-A(z,v_2)} - \frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)}\Big| \, \mathrm{d}z\Big] \\ &= \mathbf{E_1} + \mathbf{E_2}\end{aligned}$$

To study $\mathbf{E_1}$, since $f(x, v)$ is continuous and $|\sigma_{i,v_1}^2 - \sigma_{i,v_2}^2| \xrightarrow{a.s.} 0$ when $v_2 \to v_1$, one has $|f(z, \sigma_{i,v_1}^2) - f(z, \sigma_{i,v_2}^2)| \overset{a.s.}{\to} 0$. Further, since $f(x, v)$ is bounded by a constant M,

$$\left|f(z,\sigma_{i,v_1}^2)-f(z,\sigma_{i,v_2}^2)\right|\cdot\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)} \le 2M\cdot\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)} \quad \text{a.s.}$$

and the right-hand side integrates to 2M with respect to z. So by the Dominated Convergence theorem (DCT),

$$\int_{\mathbb{R}} \left|f(z,\sigma_{i,v_1}^2) - f(z,\sigma_{i,v_2}^2)\right|\cdot\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)}\,\mathrm{d}z \to 0 \quad \text{as } v_2 \to v_1.$$

This implies $\mathbf{E_1} \to 0$ as $v_2 \to v_1$.

Next we show the convergence of $\mathbf{E_2}$ by using the arguments in *Scheffé*'s theorem (cf [19], Theorem 16.12). To simplify the notation, let

$$\mathrm{d}\mu = \frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)}\,\mathrm{d}z \quad \text{and} \quad \delta_{v_2} = \frac{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}{\sqrt{2\pi\,\sigma_h^{*2}(v_2)}}\cdot e^{-A(z,v_2)+A(z,v_1)}$$

Then

$$\int_{\mathbb{R}}\Big[\frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_1)}}\cdot e^{-A(z,v_1)} - \frac{1}{\sqrt{2\pi\,\sigma_h^{*2}(v_2)}}\cdot e^{-A(z,v_2)}\Big]\,\mathrm{d}z = \int_{\mathbb{R}}\Big[\delta_{v_2}-1\Big]\,\mathrm{d}\mu.$$

Let $g_{v_2} = 1 - \delta_{v_2}$, due to the continuity of $\sqrt{2\pi\,\sigma_h^{*2}(v)}$ and $e^{-A(z,v)}$ with respect to v, we know that $g_{v_2} \overset{a.s.}{\to} 0$ almost surely when $v_2 \to v_1$. So the positive part $g_{v_2}^+$ of

g_{v_2} converges to 0 almost surely. Moreover, $0 \le g_{v_2}^+ \le 1$ and 1 is integrable with respect to $d\mu$, so the DCT applies and $\int_{\mathbb{R}} g_{v_2}^+ d\mu \to 0$. But

$$\int g_{v_2} d\mu = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\, \sigma_h^{*2}(v_1)}} \cdot e^{-A(z,v_1)} \,\mathrm{d}z - \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\, \sigma_h^{*2}(v_2)}} \cdot e^{-A(z,v_2)} \,\mathrm{d}z = 0.$$

Therefore

$$\int_{\mathbb{R}} |g_{v_2}| \,\mathrm{d}\mu = \int_{g_{v_2} \ge 0} g_{v_2} \,\mathrm{d}\mu - \int_{g_{v_2} < 0} g_{v_2} \,\mathrm{d}\mu = 2 \int_{g_{v_2} \ge 0} g_{v_2} \,\mathrm{d}\mu = 2 \int_{\mathbb{R}} g_{v_2}^+ d\mu \to 0.$$

This implies the integral in $\mathbf{E_2}$ converges to 0 as $v_2 \to v_1$. One also observes

$$\begin{aligned} &\int_{\mathbb{R}} \left| f(z, \sigma_{i,v_2}^2) \right| \cdot \Big| \frac{1}{\sqrt{2\pi\, \sigma_h^{*2}(v_2)}} \cdot e^{-A(z,v_2)} - \frac{1}{\sqrt{2\pi\, \sigma_h^{*2}(v_1)}} \cdot e^{-A(z,v_1)} \Big| \,\mathrm{d}z \\ \le &\int_{\mathbb{R}} M \Big[\frac{1}{\sqrt{2\pi\, \sigma_h^{*2}(v_2)}} \cdot e^{-A(z,v_2)} - \frac{1}{\sqrt{2\pi\, \sigma_h^{*2}(v_1)}} \cdot e^{-A(z,v_1)} \Big] \,\mathrm{d}z = 2M \end{aligned}$$

Applying the DCT again we get $\mathbf{E_2} \to 0$.

Combining the previous results we have $\mathcal{P}f(x, v_2) \to \mathcal{P}f(x, v_1)$ as $v_2 \to v_1$. And so $\mathcal{P}$ satisfies the Weak Feller property and the proof is complete. ■

Remark It should be pointed out that any Ornstein-Uhlenbeck process is weak Feller. Masuda ([52], Theorem 3.1) proved the strong Feller property for the multidimensional OU process driven by a general Lévy process. In the next section, we will use a similar approach to show the smoothness of the transition density and the Strong Feller property of $\mathcal{P}$.

2.3 The Smoothness of the Transition and the Marginal Density

In this section we will find sufficient conditions for the smoothness, that is, differentiability with respect to x and v, of the transition probability density and the (stationary) marginal density. A direct consequence of the existence of the transition density is that $\mathcal{P}$ is strong Feller, which strengthens the result in Theorem 2.2.1.

To study the smoothness of a probability measure, we need the following result:

Lemma 2.3.1: **(Sato 1999, Proposition 28.1)** Let a probability distribution function $F(\mathbf{x})$ have characteristic function $\phi(\mathbf{z})$ on $\mathbb{R}^d$ which satisfies

$$\int_{\mathbb{R}^d} |\phi(\mathbf{z})|\, |\mathbf{z}|^n \, d\mathbf{z} < \infty \tag{2.3}$$

for some $n \in \mathbb{Z}_+$. Then F has a density $f(\mathbf{x})$ of class C^n and the partial derivatives of $f(\mathbf{x})$ of orders $0, 1, \ldots, n$ tend to 0 as $|\mathbf{x}| \to \infty$.

Using a similar approach as in Masuda [52], we prove the following result:

Theorem 2.3.2: Suppose that there exist constants $\alpha \in (0, 2)$ and $c_w > 0$ such that

$$\int_{\{x:|tx|\le 1\}} (tx)^2 w(x)\, \mathrm{d}x \ \ge c_w |t|^{2-\alpha} \tag{2.4}$$

for any $t \in \mathbb{R}$ satisfying $|t| \ge 1$. Then the transition density $p(\cdot; v)$ for (X_i, σ_i^2) exists and its (partial) derivatives of all orders exist.

Proof: Due to the stationarity of (X_i, σ_i^2), for $v > 0$[1]

$$\phi(u_1, u_2; v) \;=\; E\Big[e^{iu_1X_j + iu_2\sigma_j^2}|\{\sigma_{j-1}^2 = v\}\Big] \;=\; E\Big[e^{iu_1X_1 + iu_2\sigma_1^2}|\{\sigma_0^2 = v\}\Big].$$

Therefore,

$$\begin{aligned}\phi(u_1, u_2; v) = \exp(iu_1\mu h) \cdot E\Big[&\exp\Big\{iu_1\beta\int_0^h \sigma_s^2\,\mathrm{d}s \;+ iu_1\int_0^h \sigma_s\,\mathrm{d}W_s + iu_1\rho\int_0^h \mathrm{d}Z_{\lambda s}\Big\}\\ &\cdot\exp\Big\{iu_2e^{-\lambda h}v + iu_2\int_0^h e^{-\lambda(h-s)}\,\mathrm{d}Z_{\lambda s}\Big\}\Big].\end{aligned}$$

Since W_s is independent of the $Z_{\lambda s}$, by conditioning and unconditioning on the complete trajectory of $Z_{\lambda s}$ on $s \in (0, h]$,

$$\begin{aligned}\phi(u_1, u_2; v) &= \exp(iu_1\mu h) \cdot E\Big[\exp\Big\{iu_1\beta\int_0^h \sigma_s^2\,\mathrm{d}s - \frac{u_1^2}{2}\int_0^h \sigma_s^2\,\mathrm{d}s \;+ iu_1\rho\int_0^h \mathrm{d}Z_{\lambda s}\Big\}\cdot\\ &\qquad\exp\Big\{iu_2e^{-\lambda h}v + iu_2\int_0^h e^{-\lambda(h-s)}\,\mathrm{d}Z_{\lambda s}\Big\}\Big]\\ &= \exp(iu_1\mu h) \cdot E\Big[\exp\Big\{iu_1\beta\int_0^h e^{-\lambda s}\,\mathrm{d}s\;v - \frac{u_1^2}{2}\int_0^h e^{-\lambda s}\,\mathrm{d}s\;v + iu_2e^{-\lambda h}v\Big\}\cdot\\ &\qquad\exp\Big\{\Big(-\frac{u_1^2}{2} + iu_1\beta\Big)\iint_{[0,h]\times[0,s]} e^{-\lambda(h-u)}\,\mathrm{d}Z_{\lambda u}\,\mathrm{d}s\\ &\qquad\qquad + \int_0^h \Big(iu_1\rho + iu_2e^{-\lambda(h-s)}\Big)\,\mathrm{d}Z_{\lambda s}\Big\}\Big].\end{aligned}$$

By using the fact that

$$\int_0^h\int_0^s e^{-\lambda(h-u)}\,\mathrm{d}Z_{\lambda u}\,\mathrm{d}s \;= \lambda^{-1}\int_0^h [1 - e^{-\lambda(h-s)}]\,\mathrm{d}Z_{\lambda s}\,,$$

[1]Since $\sigma_t^2 \neq 0$ with probability 1, without loss of generality, we can always assume the volatility in non-zero

we have in terms of $g(s) = \frac{1-e^{-\lambda(h-s)}}{\lambda}$,

$$\phi(u_1, u_2; v) = \exp(i u_1 \mu h) \cdot E\Big[\exp\Big\{\big(-\frac{u_1^2}{2} + i\beta u_1\big) g(0)\, v + i e^{-\lambda h} u_2 v\Big\} \cdot \exp\Big\{\big(-\frac{u_1^2}{2} + i\beta u_1\big) \int_0^h g(s)\, \mathrm{d}Z_{\lambda s} + i \int_0^h \Big(u_1 \rho + i u_2 e^{-\lambda(h-s)}\Big)\, \mathrm{d}Z_{\lambda s}\Big\}\Big].$$

Then the norm of $\phi(u_1, u_2; v)$ is given by

$$\begin{aligned} |\phi(u_1, u_2; v)| &= \exp\left\{-\frac{1-e^{-\lambda h}}{2\lambda}\, u_1^2\, v\right\} \cdot \\ &\quad \left| E\Big[\exp\Big\{\int_0^h \Big(-\frac{1-e^{-\lambda(h-s)}}{\lambda}\, \frac{u_1^2}{2} \right. \\ &\quad \left. + i\Big[\big(\rho + \frac{1-e^{-\lambda(h-s)}}{\lambda}\beta\big)u_1 + e^{-\lambda(h-s)} u_2\Big]\Big)\, \mathrm{d}Z_{\lambda s}\Big\}\Big]\right| \\ &= \exp\left\{-\frac{1-e^{-\lambda h}}{2\lambda}\, u_1^2\, v\right\} \cdot \\ &\quad \left| E\left[\exp\Big\{\int_0^h \theta(s; u_1, u_2, \lambda, h, \beta, \rho)\, \mathrm{d}Z_{\lambda s}\Big\}\right]\right| \end{aligned} \tag{2.5}$$

where

$$\theta(s; u_1, u_2, \lambda, h, \beta, \rho) = -g(s)\, \frac{u_1^2}{2} + i\Big[\big(\rho + g(s)\beta\big)u_1 + e^{-\lambda(h-s)} u_2\Big].$$

The function $g(s)$ is non-negative, decreasing and concave upward in s on $[0, h]$. To simplify the notation, we will use $\theta(s)$ instead of $\theta(s; u_1, u_2, \lambda, h, \beta, \rho)$ in the rest of the proof.

Recall the *Key Formula* in [55]: Let $f : \mathbb{R}^+ \to \mathbb{C}$ be complex and left continuous such that $\mathbf{Re}(f) \leq 0$. Then

$$E\left[\exp\left(\int_0^t f(s)\,\mathrm{d}Z_{\lambda s}\right)\right] = \exp\left(\lambda \int_0^t \kappa(f(s))\,\mathrm{d}s\right) \tag{2.6}$$

where $\kappa(\cdot)$ is the Cumulant Transform Function of Z_1.

Since $\mathbf{Re}(\theta(s)) \leq 0$ for all $s \in [0,h]$, the *Key Formula* applies. We have

$$\begin{aligned}
&\left|E\left[\exp\left\{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}\right\}\right]\right| \\
&= \left|\exp\left\{\lambda \int_0^h \kappa(\theta(s))\,\mathrm{d}s\right\}\right| \\
&= \left|\exp\left\{\lambda \int_0^h \int_{\mathbb{R}+} (e^{\theta(s)x} - 1) w(x)\ \mathrm{d}x\ \ \mathrm{d}s\right\}\right| \\
&= \left|\exp\left\{\lambda \int_0^h \int_{\mathbb{R}+} (e^{\mathbf{Re}(\theta(s))x} \cos(\mathbf{Im}(\theta(s))x) - 1) w(x)\ \mathrm{d}x\ \ \mathrm{d}s\right.\right. \\
&\qquad \left.\left. + i\lambda \int_0^h \int_{\mathbb{R}+} (e^{\mathbf{Re}(\theta(s))x} \sin(\mathbf{Im}(\theta(s))x)) w(x)\ \mathrm{d}x\ \ \mathrm{d}s\right\}\right| \\
&= \left|\exp\left\{\lambda \int_0^h \int_{\mathbb{R}+} (e^{\mathbf{Re}(\theta(s))x} \cos(\mathbf{Im}(\theta(s))x) - 1) w(x)\ \mathrm{d}x\ \ \mathrm{d}s\right\}\right| \\
&= \left|\exp\left\{\lambda \int_0^h \int_{\mathbb{R}+} \left(e^{-g(s)\frac{u_1^2}{2}x} \cos\left([(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2]\, x\right) - 1\right)\right.\right. \\
&\qquad \left. \times\, w(x)\,\mathrm{d}x\ \ \mathrm{d}s\right\}. \qquad (*)
\end{aligned}$$

In order to use Lemma 2.3.1 to prove the smoothness of the joint density, we need to verify

$$\iint\limits_{\mathbb{R}\times\mathbb{R}} |\phi(u_1, u_2; v)| \cdot |u_1^2 + u_2^2|^{k/2}\ \mathrm{d}u_1\ \mathrm{d}u_2 < \infty \qquad \text{for some} \quad k > 0,$$

and it suffices to show

$$\iint\limits_{\mathbb{R}\times\mathbb{R}} |\phi(u_1, u_2; v)| \cdot (|u_1|^k + |u_2|^k)\ \mathrm{d}u_1\ \mathrm{d}u_2\ < \infty \qquad \text{for some} \quad k > 0.$$

The main idea of the proof is the following three decompositions of the integration region:

(1) First, choose a $\Delta \in [0, h]$ such that the coefficient $(\rho + g(s)\beta)$ of u_1 in (2.5) does not change sign when s ranges within $[0, \Delta]$ or $[\Delta, h]$.

(2) Next, we wish to partition the integration over $(u_1, u_2) \in \mathbb{R}^2$ into two regions S and its complement S^c. The region S is defined in such a way that the following inequality holds for $\forall s \in [0, \Delta]$ (or $[\Delta, h]$):

$$\left|(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2\right| \geq 1 \quad \text{if} \quad (u_1, u_2) \in S.$$

The reason for this special construction is that, when finding the upper bound of $(*)$, we will encounter the integral on the left-hand side of (2.4) with t replaced by $|(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2|$. By restricting (u_1, u_2) in S, we can use condition (2.4) on x. Meanwhile, the above construction indeed gives linear bounds on u_2 in terms of u_1 when $(u_1, u_2) \in S^c$. Depending on the signs of the parameter, one may have different bounds for u_2, without further constraints on the parameters, let us just denote those linear functions as l_1, l_2, l_3 and l_4.

Since $\left|E\left[\exp\left\{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}\right\}\right]\right|$ is trivially bounded by 1, one has

$$\begin{aligned}
&\iint\limits_{S^c} |\phi(u_1, u_2; v)| \cdot (|u_1|^k + |u_2|^k)\ \mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad \le \iint\limits_{S^c} \exp\Big\{-\frac{v(1-e^{-\lambda h})}{2\lambda}\, u_1^2\Big\}\, (|u_1|^k + |u_2|^k)\, \mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad = \Big(\int\limits_{\mathbb{R}^-}\int\limits_{l_1}^{l_2} + + \int\limits_{\mathbb{R}^+}\int\limits_{l_3}^{l_4}\Big)\ \exp\Big\{-\frac{v(1-e^{\lambda h})}{2\lambda}\, u_1^2\Big\}(|u_1|^k + |u_2|^k)\, \mathrm{d}u_2\ \mathrm{d}u_1\ < \infty.
\end{aligned}$$

So we need only to focus on the integration over S. The explicit forms of the l_i's will be given later in the proof.

(3) Once $\varDelta$ and S are given, for every $(u_1, u_2) \in S$, define another region $S_X \subset \mathbb{R}^+$ by

$$\begin{aligned}
S_X &\triangleq S_X(\varDelta, u_1, u_2) \\
&= \Big\{x : \big|x[(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2]\big| \le \frac{\pi}{2} \\
&\qquad \text{for } (u_1, u_2) \in S \quad \text{and} \quad s \in (0, \varDelta]\Big\}
\end{aligned}$$

Notice the integrand in $(*)$ is non-positive, so we can bound the integral of x over $\mathbb{R}^+$ by the one over S_X. Using the inequality $1-\cos(x) \ge 2(\frac{x}{\pi})^2$ for $|x| \le \pi$ and condition (2.4), we find the desired upper bound for $(*)$.

Next we give details on how to construct these partitions and prove the integrability. Since there are unknown parameters β and ρ in the coefficient of u_1, to avoid adding more complexity to the already involved notations, we will proceed in the proof by separate consideration of three mutually exclusive and exhaustive cases:

- **Case 1:** $\rho < 0$;
- **Case 2:** $\rho = 0$ and $\beta \neq 0$;
- **Case 3:** $\rho = 0$ and $\beta = 0$.

Case 1 First, we study the sign of $(\rho + g(s)\beta)$:

(1) $\beta < 0$, then $\rho + g(s)\beta < 0$ for all $s \in [0, h]$.

(2) $\beta > 0$ and $1 + \lambda\rho/\beta < 0$, then $\rho + g(s)\beta < 0$ for all $s \in [0, h]$.

(3) $\beta > 0$ but $1 + \lambda\rho/\beta \geq 0$, then $\rho + g(s)\beta < 0$ for all $s > h + \lambda^{-1}\ln(1 + \lambda\rho/\beta)$.

One observes that by choosing $\Delta = h + \lambda^{-1}\ln(1 + \lambda\rho/\beta)$, $\rho + g(s)\beta < 0$ for $\forall s \in (\Delta, h]$. Now fix this Δ and define S and S_X respectively for (u_1, u_2) and x by:

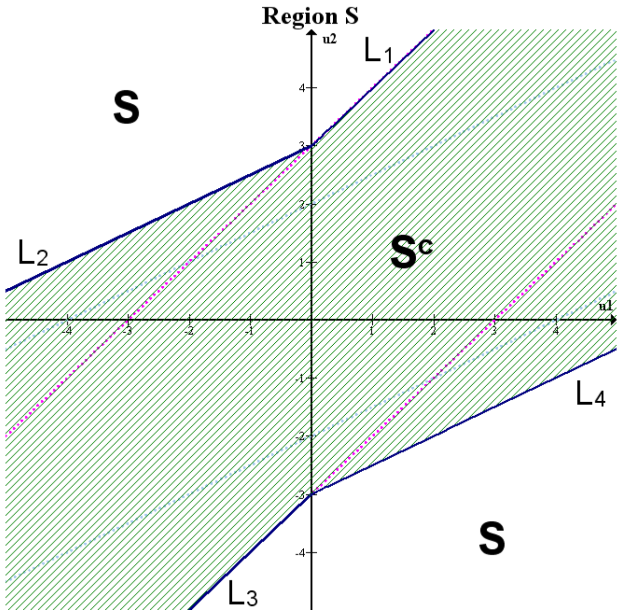
$$S \triangleq \Big\{(u_1, u_2) : u_2 \geq -\rho u_1 + e^{\lambda h} \text{ or } u_2 \leq -e^{\lambda h}(\rho + g(\Delta)\beta)u_1 - e^{\lambda h} \text{ when } u_1 \leq 0;$$
$$u_2 \geq -e^{\lambda h}(\rho + g(\Delta)\beta)u_1 + e^{\lambda h} \text{ or } u_2 \leq -\rho u_1 - e^{\lambda h} \text{ when } u_1 \geq 0.\Big\}.$$

and

$$S_X \triangleq \Big\{x \,:\, |(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2| \cdot x \leq \frac{\pi}{2}$$
$$\text{where } x \in \mathbb{R}^+, (u_1, u_2) \in S, s \in (\Delta, h]\Big\}$$

One can verify that $|(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2| \geq 1$ for $(u_1, u_2) \in S$.

The following figure explains the idea of S:



Line L_1:
$u_2 = -(\rho + g(\Delta)\beta)e^{\lambda h}u_1 + e^{\lambda h}$

Line L_2:
$u_2 = -\rho u_1 + e^{\lambda h}$

Line L_3:
$u_2 = -(\rho + g(\Delta)\beta)e^{\lambda h}u_1 - e^{\lambda h}$

Line L_4:
$u_2 = -\rho u_1 - e^{\lambda h}$

When $x \in S_X$ and $s \in (\Delta, h]$, the integral in $(*)$ with respect to x becomes:

$$\begin{aligned}
&\int_{\mathbb{R}^+} \left(e^{-g(s)\,\frac{u_1^2}{2}\,x} \cos\left([(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2]\,x \right) - 1 \right) w(x)\ \mathrm{d}x \\
\leq &\int_{S_X} \left(e^{-g(s)\,\frac{u_1^2}{2}\,x} \cos\left([(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2]\,x \right) - 1 \right) w(x)\ \mathrm{d}x \\
\leq &\int_{S_X} \left(\cos\left([(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2]\,x \right) - 1 \right) w(x)\ \mathrm{d}x \\
\leq - &\int_{S_X} 2\frac{[(\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2]^2 x^2}{\pi^2} w(x)\ \mathrm{d}x \\
\leq - &\frac{2c_w}{\pi^{2-\alpha}} \left| (\rho + g(s)\beta)u_1 + e^{-\lambda(h-s)}u_2 \right|^{2-\alpha}
\end{aligned}$$

The first inequality follows because the integrand is non-positive and $S_X \subset \mathbb{R}^+$ and the second holds because the cosine term is non-negative on S_X. The third inequality uses the inequality $1 - \cos x \geq 2\left(\frac{x}{\pi}\right)^2$ for $|x| \leq \pi$. The last line holds under the assumed condition (2.4).

We can rewrite the term $|(\rho+g(s)\beta)u_1+e^{-\lambda(h-s)}u_2|$ in the following way:

$$\left|(\rho+g(s)\beta)u_1+e^{-\lambda(h-s)}u_2\right| = \left|u_1\big(\rho+\frac{\beta}{\lambda}\big)+\big(-\frac{\beta}{\lambda}u_1+u_2\big)e^{-\lambda h}\cdot e^{\lambda s}\right|$$

When $(u_1,u_2)\in S$, $|u_1(\rho+g(s)\beta)+u_2e^{-\lambda(h-s)}|>1$, so the term in the absolute value does not change sign as s varies in $(\Delta,h]$. Further, as $e^{-\lambda s}$ is a monotone function of s for any fixed value of (u_1,u_2), $|u_1(\rho+g(s)\beta)+u_2e^{-\lambda(h-s)}|$ must obtain its minimum either at Δ or h. That is, for any fix (u_1,u_2):

$$\begin{aligned}&\left|u_1(\rho+g(s)\beta)+u_2e^{-\lambda(h-s)}\right|^{2-\alpha}\\ \geq\ &\min\left(\left|u_1(\rho+g(\Delta)\beta)+u_2e^{-\lambda(h)}\right|^{2-\alpha}, |u_1(\rho+g(h)\beta)+u_2|^{2-\alpha}\right)\\ \triangleq\ &|c_1u_1+c_2u_2|^{2-\alpha}\end{aligned}$$

for $\forall s\in(\Delta,h]$ with non-zero $c_1=\rho+g(\Delta)\beta,\ c_2=e^{-\lambda(h)}$ (or $c_1=\rho+g(h)\beta,\ c_2=1$).

Now we have an explicit bound on $|E[\exp\{\int_0^h\theta(s)dZ_{\lambda s}\}]|$ on S:

$$\left|E\left[\exp\left\{\int_0^h\theta(s)\,\mathrm{d}Z_{\lambda s}\right\}\right]\right|\leq\exp\left\{-K\left|c_1u_1+c_2u_2\right|^{2-\alpha}\right\},$$

where $K=\frac{2\lambda hc_w}{\pi^{2-\alpha}}$.

At last, we are ready to show $\iint\limits_{\mathbb{R}\times\mathbb{R}}|\phi(u_1,u_2;v)|\cdot(|u_1|^k+|u_2|^k)\,\mathrm{d}u_1\,\mathrm{d}u_2$ is finite. Recall the following decomposition shown earlier in the proof:

$$
\begin{aligned}
&\iint\limits_{\mathbb{R}\times\mathbb{R}} |\phi(u_1,u_2;v)|\cdot(|u_1|^k+|u_2|^k)\ \mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad= \Big(\iint\limits_{S} + \iint\limits_{S^c}\Big)\exp\Big\{-\frac{v(1-e^{-\lambda h})}{2\lambda}\,u_1^2\Big\}\cdot\Big|E[e^{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}}]\Big|(|u_1|^k+|u_2|^k)\,\mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad\le \iint\limits_{S}\exp\Big\{-\frac{v(1-e^{-\lambda h})}{2\lambda}\,u_1^2\Big\}\cdot\exp\Big\{-K\big|c_1u_1+c_2u_2\big|^{2-\alpha}\Big\}(|u_1|^k+|u_2|^k)\,\mathrm{d}u_1\ \mathrm{d}u_2 \\
&\qquad+\iint\limits_{S^c}\exp\Big\{-\frac{v(1-e^{-\lambda h})}{2\lambda}\,u_1^2\Big\}(|u_1|^k+|u_2|^k)\,\mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad= \mathbf{i}+\mathbf{ii}
\end{aligned}
$$

Integral **i** is clearly finite since $v>0$. For integral **ii**, $|E[e^{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}}]|$ has a trivial upper bound 1, thus

$$
\begin{aligned}
\mathbf{ii} &\le \int_{\mathbb{R}^-}\int_{-e^{\lambda h}(\rho+g(\Delta)\beta)u_1-e^{\lambda h}}^{-\rho u_1+e^{\lambda h}}\exp\Big\{-\frac{v(1-e^{\lambda h})}{2\lambda}\,u_1^2\Big\}(|u_1|^k+|u_2|^k)\,\mathrm{d}u_2\ \mathrm{d}u_1 \\
&\quad+\int_{\mathbb{R}^+}\int_{-\rho u_1-e^{\lambda h}}^{-e^{\lambda h}(\rho+g(\Delta)\beta)u_1+e^{\lambda h}}\exp\Big\{-\frac{v(1-e^{\lambda h})}{2\lambda}\,u_1^2\Big\}(|u_1|^k+|u_2|^k)\,\mathrm{d}u_2\ \mathrm{d}u_1 \\
&< \infty
\end{aligned}
$$

We can conclude that under condition (2.4),

$$
\iint\limits_{\mathbb{R}\times\mathbb{R}} |\phi(u_1,u_2;v)|\cdot|u_1^2+u_2^2|^{k/2}\ \mathrm{d}u_1\ \mathrm{d}u_2\ <\infty
$$

for any positive integer k. So the transition density $p(y_1,y_2;v)$ is infinitely many times differentiable in both arguments.

Case 2 If $\rho = 0$ but $\beta \neq 0$, then $(*)$ becomes:

$$\begin{aligned}
&\left| E\left[\exp\left\{ \int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s} \right\} \right] \right| \\
&= \left| \exp\left\{ \lambda \int_0^h \int_{\mathbb{R}^+} \left(e^{-g(s)\frac{u_1^2}{2}x} \cos([g(s)\beta u_1 + e^{-\lambda(h-s)}u_2]\,x) - 1 \right) w(x)\,\mathrm{d}x\,\mathrm{d}s \right\} \right| \\
&< \left| \exp\left\{ \lambda \int_0^\Delta \int_{\mathbb{R}^+} \left(e^{-g(s)\frac{u_1^2}{2}x} \cos([g(s)\beta u_1 + e^{-\lambda(h-s)}u_2]\,x) - 1 \right) w(x)\,\mathrm{d}x\,\mathrm{d}s \right\} \right|
\end{aligned}$$

And the coefficient of u_1 becomes $g(s)\beta$. Evidently the sign of β won't affect the final conclusion since we only require the sign of $g(s)\beta$ remains unchanged. By assuming $\beta < 0$ and choosing Δ to be strictly less than h (to avoid zero coefficient of u_1), define S and S_X respectively for (u_1, u_2) and x as follows:

$$\begin{aligned}
S \triangleq \Big\{ (u_1, u_2) : \; & u_2 \geq -g(\Delta)\beta e^{\lambda(h-\Delta)}u_1 + e^{\lambda h} \\
& \qquad \text{or } \; u_2 \leq -g(0)\beta e^{\lambda h}u_1 - e^{\lambda h} \; \text{ when } \; u_1 \leq 0; \\
& u_2 \geq -g(0)\beta e^{\lambda h}u_1 + e^{\lambda h} \\
& \qquad \text{or } \; u_2 \leq -g(\Delta)\beta e^{\lambda(h-\Delta)}u_1 - e^{\lambda h} \; \text{ when } \; u_1 \geq 0. \Big\}
\end{aligned}$$

and

$$S_X \triangleq \left\{ x \,:\, |g(s)\beta u_1 + e^{-\lambda(h-s)}u_2| \cdot x \leq \frac{\pi}{2} \;\text{ where }\; x \in \mathbb{R}^+, (u_1, u_2) \in S, s \in [0, \Delta] \right\}.$$

When $(u_1, u_2) \in S$, we have $|g(s)\beta u_1 + e^{-\lambda(h-s)}u_2| \geq 1$. Following the similar arguments in **Case 1**, one can show that for all $(u_1, u_2) \in S$ and $s \in [0, \Delta]$:

$$\int_{\mathbb{R}+} \left(e^{-g(s)\frac{u_1^2}{2}x} \cos\left([g(s)\beta u_1 + e^{-\lambda(h-s)} u_2]\, x \right) - 1 \right) w(x)\, \mathrm{d}x$$
$$\leq -\frac{2c_w}{\pi^{2-\alpha}} \left| g(s)\beta u_1 + e^{-\lambda(h-s)} u_2 \right|^{2-\alpha}$$

Then the rest of the proof proceeds in the same way as **Case 1** with the new region S.

Case 3 If $\rho = 0$ and $\beta = 0$, then

$$\left| E\left[\exp\left\{ \int_0^h \theta(s)\, \mathrm{d}Z_{\lambda s} \right\} \right] \right|$$
$$< \left| \exp\left\{ \lambda \int_0^h \int_{\mathbb{R}+} \left(e^{-g(s)\frac{u_1^2}{2}x} \cos(e^{-\lambda(h-s)} u_2\, x) - 1 \right) w(x)\, \mathrm{d}x\, \mathrm{d}s \right\} \right| .$$

In this case there is no need to choose any Δ. The the region S simplified to $\{u_2 : |u_2| > e^{\lambda(h-s)}\}$ and $S_X = \{x : |e^{-\lambda(h-s)} u_2| x \leq \frac{\pi}{2}$ for $u_2 \in S\,, s \in [0, h]\}$. Then follow the arguments in **Case 1** and use (2.4), one can verify the integrability.

To summarize, in all three cases of parameter specifications, $p(y_1, y_2; v)$ is infinitely differentiable under the given conditions. ■

Next we study the strong Feller property of $\mathcal{P}$. Recall that $\mathcal{P}$ is called *strong Feller* if

$$\mathcal{P} f \in C_b(\mathcal{S}) \quad \text{for any} \quad f \in b\mathcal{B}(\mathcal{S}). \tag{2.7}$$

That is, $\mathcal{P}$ maps a bounded $\mathcal{S}$-measurable function to a continuous bounded $\mathcal{S}$-

measurable function. We need the following proposition for our proof.

Proposition 2.3.3: (**Parseval and Plancherel,** [**77**]) Let $f(\mathbf{t})$ and $g(\mathbf{t})$ be the characteristic functions of two absolutely continuous distributions with density $p(\mathbf{x})$ and $q(\mathbf{x})$ respectively, then

$$\int_{\mathbb{R}^m} [p(\mathbf{x}) - q(\mathbf{x})]^2 d\mathbf{x} = \frac{1}{(2\pi)^m} \int_{\mathbb{R}^m} |f(\mathbf{t}) - g(\mathbf{t})|^2 d\mathbf{t}$$

provided that the integrals exist.

Lemma 2.3.4: Under condition (2.4), the transition operator $\mathcal{P}$ for (X_i, σ_i^2) is Strong Feller.

Proof: Let $\phi(u_1, u_2; v)$ be the characteristic function of $P(\cdot; v)$. From Theorem 2.3.2 we know that under condition (2.4) one has $\int_{\mathbb{R}^2} |\phi(u_1, u_2; v)| \, \mathrm{d}u_1 \, \mathrm{d}u_2 < \infty$, which implies that the transition density $p(\cdot; v)$ exists. In fact, one can also show that $\int_{\mathbb{R}^2} |\phi(u_1, u_2; v)|^2 \, \mathrm{d}u_1 \, \mathrm{d}u_2 < \infty$ as we have an exponential bound on $|\phi(u_1, u_2; v)|$. This implies $p(\cdot; v) \in L^2$ and we can use Proposition (2.3.3) to prove the convergence of $p(y_1, y_2; v_2)$ to $p(y_1, y_2; v_2)$ when $v_2 \to v_1$.

Since the 1-step transition density $p(y_1, y_2; v)$ exists, to prove (2.7) it is equivalent to prove

$$\Big| \int_{\mathbb{R}} \int_{\mathbb{R}+} f(y_1, y_2) p(y_1, y_2; v_1) \, \mathrm{d}y_1 \, \mathrm{d}y_2 - \int_{\mathbb{R}} \int_{\mathbb{R}+} f(y_1, y_2) p(y_1, y_2; v_2) \, \mathrm{d}y_1 \, \mathrm{d}y_2 \Big| \to 0.$$

as $v_2 \to v_1$. By the boundedness of f, one needs to show

$$\int_{\mathbb{R}} \int_{\mathbb{R}^+} |p(y_1, y_2; v_1) - p(y_1, y_2; v_2)| \, \mathrm{d}y_1 \, \mathrm{d}y_2 \to 0,$$

We want to use the *Scheffé*'s theorem again to show the above convergence. First to show that

$$p(y_1, y_2; v_2) \to p(y_1, y_2; v_1) \qquad \text{as} \qquad v_2 \to v_1.$$

When $v_2 \in (v_1 - \epsilon, v_1 + \epsilon)$, using the characteristic function expression (2.5):

$$\begin{aligned}
&|\phi(u_1, u_2; v_1) - \phi(u_1, u_2; v_2)|^2 \\
&\qquad \leq 2\Big[|\phi(u_1, u_2; v_1)|^2 + |\phi(u_1, u_2; v_2)|^2\Big] \\
&\qquad = 2\,\Big|E\big[\exp\{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}\}\big]\Big|^2 \cdot \\
&\qquad\qquad \Big[\exp\Big\{-\frac{(1-e^{-\lambda h})}{\lambda} u_1^2\, v_1\Big\} + \exp\Big\{-\frac{(1-e^{-\lambda h})}{\lambda} u_1^2\, v_2\Big\}\Big] \\
&\qquad \leq 4\,\exp\Big\{-\frac{(1-e^{-\lambda h})}{\lambda} u_1^2\,(v_1 - \epsilon)\Big\}\Big|E\big[\exp\{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}\}\big]\Big|^2
\end{aligned}$$

The last term is integrable following the proof of Theorem 2.3.2. Thus $|\phi(u_1, u_2; v_1) - \phi(u_1, u_2; v_2)|^2$ is bounded by an integrable function which depends on v_1 only. Further observe that

$$\begin{aligned}
&|\phi(u_1, u_2; v_1) - \phi(u_1, u_2; v_2)|^2 \\
&\quad = \Big[\exp\Big\{-\frac{(1-e^{-\lambda h})}{\lambda} u_1^2\, v_1\Big\} - \exp\Big\{-\frac{(1-e^{-\lambda h})}{\lambda} u_1^2\, v_2\Big\}\Big] \cdot \Big|E\big[\exp\{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}\}\big]\Big|^2 \\
&\quad \to 0
\end{aligned}$$

as $v_2 \to v_1$. By the Dominated Convergence theorem and Proposition 2.3.3

$$\begin{aligned}
&\int_{\mathbb{R}} \int_{\mathbb{R}+} \left| p(y_1, y_2; v_1) - p(y_1, y_2; v_2) \right|^2 dy_1 dy_2 \\
&= \frac{1}{4\pi^2} \int_{\mathbb{R}^2} \left| \phi(u_1, u_2; v_1) - \phi(u_1, u_2; v_2) \right|^2 \mathrm{d}u_1 \, \mathrm{d}u_2 \to 0.
\end{aligned}$$

To derive the convergence of $|p(y_1, y_2; v_1) - p(y_1, y_2; v_2)|$ to 0 when $v_2 \to v_1$, consider the L_2 convergence above along the sequence $\{v_{2,1}, v_{2,2}, \ldots, v_{2,n}\}$. Using the argument in ([46], Pg. 292), one can find a subsequence $\{v_{2,n_1}, v_{2,n_2}, \ldots, v_{2,n_k}\}$ where $p(y_1, y_2; v_{2,n_k}) \to p(y_1, y_2; v_1)$ as $k \to \infty$. As the L_2 space is a complete metrizable space, the convergence of along the subsequence is the same as the convergence as in the original sequence.

Using *Scheffé*'s theorem as we did in the proof of the weak Feller property, one finds

$$\int_{\mathbb{R}} \int_{\mathbb{R}+} |p(y_1, y_2; v_1) - p(y_1, y_2; v_2)| dy_1 dy_2 \to 0$$

as $v_2 \to v_1$. Therefore, $\mathcal{P}$ is strong Feller. ■

Remark:

(1) When Z_t is a univariate subordinator, condition (2.4) is simplified to:

$$\int_0^{1/|v|} (vx)^2 w(x) \, \mathrm{d}x \quad \geq \quad c_w |v|^{2-\alpha} \quad \Leftrightarrow \quad \int_0^{1/|v|} x^2 \, w(x) \, \mathrm{d}x \quad \geq \quad c_w |v|^{-\alpha}$$

for $|v| > 1$ and $\alpha \in (0, 2)$. This condition in fact requires the Lévy process to have high level activity for small jumps. That is, $w(x)$ needs to behave like x^{-k} for $k \in (1, 3)$ when x is close to 0. This implies the pure jump process Z_t

has infinite many jumps (activities) in any finite time interval. Furthermore, if $k > 2$, then Z_t has infinite variation[2]. A nonparametric study conducted by Todorov and Tauchen [72] suggests the activity level of the VIX is substantially higher than a finite activity process. This fact justifies the condition (2.4) as more than a technical convenience.

(2) It turns out that the Γ-OU process does not satisfy condition (2.4) while the IG-OU process and Tempered-Stable-OU processes do. The reason is that the BDLP for the Γ-OU process is not infinitely active on any finite time horizon. Without condition (2.4) it will be hard to prove the smoothness of the joint transition density, but still we can prove the smoothness of joint density of (X_i, σ_i^2) thanks to the explicit characteristic function of σ_0^2.

Theorem 2.3.5: Assuming that condition (2.4) holds, then the joint (stationary) distribution $F(x, v)$ of (X_j, σ_j^2) has partial derivatives of all orders.

Proof: Let $\phi(u_1, u_2)$ be the characteristic function of $F(x, v)$. We want to show the following is true for all positive k:

$$\iint |\phi(u_1, u_2)|\, |u_1^2 + u_2^2|^{k/2} \,\mathrm{d}u_1 \ \mathrm{d}u_2 < \infty.$$

In the study of the smoothness of the transition density, we derive the characteristic function for the transition density $p(y_1, y_2; v)$. Following the same steps and by

[2] $k > 2$ is not in the scope of BN-S model since technically a subordinator doesn't have infinite variation.

recognizing that σ_0^2 is independent of $(W_s, Z_{\lambda s})$ for $s \in (0, h]$, one can derive the characteristic function $\phi(u_1, u_2)$ for $F(x, v)$ and get a similar upper bound:

$$
\begin{aligned}
&|\phi(u_1, u_2)| \\
&\qquad \le \left| E\left[\exp\left\{\left[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}\right]\sigma_0^2\right\}\right]\right| \cdot \left| E\left[\exp\left\{\int_0^h \theta(s) dZ_{\lambda s}\right\}\right]\right| \\
&\qquad \le \left| E\left[\exp\left\{\left[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}\right]\sigma_0^2\right\}\right]\right| \cdot \\
&\qquad\qquad \left(e^{-C(u_1,u_2)}\mathbb{I}_{(u_1,u_2)\in S} + |E[e^{\int_0^h \theta(s)\,\mathrm{d}Z_{\lambda s}} \cdot \mathbb{I}_{(u_1,u_2)\in S^c}]|\right)
\end{aligned}
$$

where $C(u_1, u_2) = C_1|c_1 u_1 + c_2 u_2|^{2-\alpha}$ and $g(0)$, $\theta(s)$ and region S are defined exactly the same as in the proof of Theorem 2.3.2. Since $e^{-C(u_1,u_2)}$ dominates the polynomials of u_1 and u_2 of all orders, so we need only to consider the finiteness of the integral:

$$
\iint_{S^c} \left| E\left[\exp\left\{[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}]\sigma_0^2\right\}\right]\right| \cdot (|u_1|^k + |u_2|^k)\,\mathrm{d}u_2\ \mathrm{d}u_1
$$

Recall the upper and lower bound on u_2 in S^c are all linear in u_1, it suffices to check

$$
\int_{\mathbb{R}} \left| E\left[\exp\left\{[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}]\sigma_0^2\right\}\right]\right| \cdot |u_1|^{k+1}\,\mathrm{d}u_1 < \infty.
$$

Choose an C_2 such that $C_2^2 g(0)/2 > 1$. Considering the expected value term in the integrand for $|u| > C_2$,

$$
\begin{aligned}
&\left|E\left[\exp\left\{[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}]\sigma_0^2\right\}\right]\right| \\
&\qquad \le E\left[\left|\exp\left\{[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}]\sigma_0^2\right\}\right|\right] \\
&\qquad = E\left[\exp\left\{[-g(0)\frac{u_1^2}{2}]\sigma_0^2\right\}\right] \\
&\qquad = \exp\left\{\int_0^\infty \int_{\mathbb{R}+} \left(e^{-g(0)e^{-s}\frac{u_1^2}{2}x} - 1\right) w(x)\,\mathrm{d}x\ ds\right\} \\
&\qquad \le \exp\left\{\int_0^\Delta \int_{\{x:g(0)\frac{u_1^2}{2}x<1\}} \left(e^{-g(0)e^{-s}\frac{u_1^2}{2}x} - 1\right) w(x)\,\mathrm{d}x\ ds\right\} \\
&\qquad \le \exp\left\{\int_0^\infty \int_{\{x:g(0)\frac{u_1^2}{2}x<1\}} -\frac{1}{4}\left(g(0)e^{-s}\frac{u_1^2}{2}x\right) w(x)\,\mathrm{d}x\ ds\right\}. \qquad (\Delta)
\end{aligned}
$$

The third equality uses the fact that $\sigma_0^2 \stackrel{\mathcal{D}}{=} \int_0^\infty e^{-s} dZ_s$ and the last inequality holds since $e^x - 1 < x/4$ when $-1 < x < 0$. Since $(g(0)\frac{u_1^2}{2}x) < 1$ in the last integrand, by condition (2.4),

$$
\begin{aligned}
\int\limits_{\{x:g(0)\frac{u_1^2}{2}x<1\}} \left(g(0)\frac{u_1^2}{2}x\right) w(x)\,\mathrm{d}x &\ge \int\limits_{\{x:g(0)\frac{u_1^2}{2}x<1\}} \left(g(0)\frac{u_1^2}{2}x\right)^2 w(x)\,\mathrm{d}x \\
&\ge \tilde{C}_w \left|g(0)\frac{u_1^2}{2}\right|^{2-\alpha} \triangleq C_3 |u_1|^{4-2\alpha}
\end{aligned}
$$

Then Δ is bounded by:

$$\Delta \le \exp\left(-\frac{C_3}{4}|u_1|^{4-2\alpha}\right)$$

for sufficiently large u_1. Therefore,

$$\int_{\mathbb{R}} \left|E\left[\exp\left\{\left[(-\frac{u_1^2}{2} + i\beta u_1)g(0) + iu_2 e^{-\lambda h}\right]\sigma_0^2\right\}\right]\right| \cdot |u_1|^{k+1}\,\mathrm{d}u_1 < \infty.$$

for any positive k. This implies the joint distribution F has partial derivatives of all orders.

Lemma 2.3.6: If σ_t^2 is a $\Gamma(\alpha,\nu)$-OU process then the joint density $f(x,v)$ of (X_i,σ_i^2) has all k-th order partial derivatives if $\nu > k+1$.

Proof: Since in the Γ-OU case the process σ_t^2 is stationary and has a Gamma(ν,α) distribution, we can use the explicit characteristic function of the joint distribution $F(x,v)$ to study its smoothness. Similar to the proof of Theorem 2.3.5:

$$
\begin{aligned}
&|\phi(u_1,u_2)| \\
&\le \left|E\left[\exp\left\{[(-\frac{u_1^2}{2}+i\beta u_1)\frac{e^{-\lambda h}}{\lambda}+iu_2e^{-\lambda h}]\sigma_0^2\right\}\right]\right|\cdot\left|E\left[\exp\left\{\int_0^h\theta(s)\,\mathrm{d}Z_{\lambda s}\right\}\right]\right| \\
&\le \left|E\left[\exp\left\{[(-\frac{u_1^2}{2}+i\beta u_1)\frac{e^{-\lambda h}}{\lambda}+iu_2e^{-\lambda h}]\sigma_0^2\right\}\right]\right|
\end{aligned}
$$

with

$$\theta(s) = -\frac{u_1^2}{2}g(s)+i[u_1(\rho+g(s)\beta)+u_2e^{-\lambda(h-s)}]$$

and $g(s) = \left(1-e^{-\lambda(h-s)}\right)/\lambda$. The second inequality holds because $\mathbf{Re}(\theta(s)) < 0$ so the norm is less than 1. Since the marginal distribution of σ_t^2 is Gamma(α,ν), the Laplace transform of the Gamma(α,ν) density function is given by

$$E[e^{\theta\sigma_0^2}] = \left(\frac{\alpha}{\alpha-\theta}\right)^{\nu}$$

for any θ such that $\mathbf{Re}(\theta) < \alpha$. Due to the fact that $\mathbf{Re}((-\frac{u_1^2}{2}+i\beta u_1)\frac{e^{-\lambda h}}{\lambda}+$

$iu_2e^{-\lambda h}) < 0 < \alpha$,

$$E\left[\exp\left\{[(-\frac{u_1^2}{2}+i\beta u_1)\frac{e^{-\lambda h}}{\lambda}+iu_2e^{-\lambda h}]\sigma_0^2\right\}\right] = \left[\frac{\alpha}{\alpha+(\frac{u_1^2}{2}-i\beta u_1)\frac{e^{-\lambda h}}{\lambda}-iu_2e^{-\lambda h}}\right]^{\nu}.$$

Consider a subset $S \subset \mathbb{R}^{+2}$ where both $|u_1|$ and $|u_2|$ are greater than some sufficiently large positive number C_4, we have for $k \geq 1$:

$$\begin{aligned}
&\iint\limits_S |\phi(u_1,u_2)| \cdot |u_1^2+u_2^2|^{k/2}\ \mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad \leq \tilde{C}_k \iint\limits_S \left|\frac{\alpha}{\alpha+(\frac{u_1^2}{2}-i\beta u_1)\frac{e^{-\lambda h}}{\lambda}-iu_2e^{-\lambda h}}\right|^{\nu} \cdot \Big(|u_1|^k+|u_2|^k\Big)\ \mathrm{d}u_1\ \mathrm{d}u_2 \\
&\quad < \int_{|u_1|\geq C_4}\int_{|u_2|\geq C_4} \frac{C_k\,\alpha^{\nu}}{\left[\alpha^2+\frac{e^{-2\lambda h}}{4\lambda^2}u_1^4+(\frac{\beta e^{\lambda h}}{\lambda}u_1+e^{-\lambda h}u_2)^2\right]^{\nu/2}} \cdot \Big(|u_1|^k+|u_2|^k\Big)\ \mathrm{d}u_1\ \mathrm{d}u_2\,.
\end{aligned}$$

It is clear that when $\nu > k+1$, the above integral is finite, then the joint density $f(x,v)$ is k times differentiable. ■

Remark To establish the smoothness property of the transition probability distribution does not seem to be easy without the use of characteristic function. The proof will be left for future research and will not be pursued further in this paper.

2.4 Geometric Ergodicity of (X_i, σ_i^2)

Here we list all the definitions and terminologies to be used in this section. More details can be found in [53] and [54]. In Appendix **A**, we include four related lemmas and the proof of one lemma for the reader's reference

1. **α-mixing and β-mixing:** The notions of mixing are related to measuring the dependence between σ-fields. The mixing concept is particularly useful when studying the consistency and asymptotic normality of statistics when the underlying data is dependent. There are various notions of mixing and we only focus on two of them. Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and $\mathcal{U}$, $\mathcal{V}$ be two sub $-\,\sigma$-algebras of $\mathcal{F}$.

   (a) **α-mixing coefficient**:

   $$\alpha(\mathcal{U}, \mathcal{V}) = \mathrm{Sup}\Big\{ |\mathrm{P}(U)\mathrm{P}(V) - \mathrm{P}(U \cap V)|; U \in \mathcal{U},\ V \in \mathcal{V}\Big\}$$

   α-mixing is also called *strong mixing*. If the σ-algebras are generated by a stochastic process X_t, that is, $\mathcal{F}_t = \mathcal{N} \cup \sigma\{X_s, s \le t\}$, and $\mathcal{U}$ and $\mathcal{V}$ are "separated" by k time units, that is, $\mathcal{U} = \sigma\{X_s, s \le t\}$ and $\mathcal{V} = \sigma\{X_s, s \ge t + k\}$, then $\alpha(\mathcal{U}, \mathcal{V})$ is also denoted as $\alpha_X(k)$.

   (b) **β-mixing coefficient**:

   $$\beta(\mathcal{U}, \mathcal{V}) = E\text{ess-sup}\{|\mathrm{P}(V|\mathcal{U}) - \mathrm{P}(V)|; V \in \mathcal{V}\}.$$

   If the σ-algebras are generated by a Markov process X_t with limiting distribution F and transition probability $P_t(\cdot, x)$, then the β-mixing coefficient $\beta_X(t)$ is defined as:

   $$\begin{aligned} \beta_X(t) &\triangleq \int \| P_t(\cdot, x) - F(\cdot) \|_{TV}\ F(\mathrm{d}x) \\ &= \int \sup_{|f| \le 1} |P_t f(x) - F(f)| F(\mathrm{d}x) \end{aligned}$$

where $F(f) = \int f(y)dy$ and $\|m\|_{TV}$ is the *Total Variation Norm* defined by:

$$\|m\|_{TV} \triangleq \sup_{f:|f|\le 1} |m(f)| = \sup_{A\in\mathcal{B}(\mathbf{S})} m(A) - \inf_{A\in\mathcal{B}(\mathbf{S})} m(A).$$

for signed measure m on $\mathcal{B}(\mathcal{S})$.

2. **Δ-skeleton chain:** Let X^{Δ} be defined as the discrete-time Markov chain regularly sampled from X_t at time points $0, \Delta, 2\Delta, \dots$ for a constant $\Delta > 0$. We call $X^{\Delta} := \left(X_n^{\Delta}\right)_{n\in\mathbb{N}_0}$ with $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ the *Δ-skeleton chain.*

3. **φ-irreducible:** For a σ-finite measure φ on $\mathcal{B}(\mathcal{S})$, a discrete time Markov chain X^{Δ} is called *φ-irreducible* if $\sum_{n=1}^{\infty} P_{n\Delta}(A, x) > 0$ for any $x \in \mathcal{S}$ and $A \in \mathcal{B}(\mathcal{S})$ such that $\varphi(A) > 0$. We shall omit the Δ when there is no confusion.

4. **Simultaneously φ-irreducible:** Let $(\mathcal{P}_t)_{t\in\mathbb{R}_+}$ be the transition semigroup generated by X_t. Then $\mathcal{P}_t$ is *simultaneously φ-irreducible* (for some σ-finite measure φ) if all the associated Δ-skeleton chains X^{Δ} are φ-irreducible.

5. **Small Set** A set $C \in \mathcal{B}(X^{\Delta})$ is called a *small set* if there exists an $n > 0$ and a non-trivial measure ν_n on $\mathcal{B}(X^{\Delta})$ such that for all $x \in C$, $B \in \mathcal{B}(X^{\Delta})$,

$$P_n(B, x) \ge \nu_n(B)$$

When the above inequality holds, we also say C is ν_n-small.

6. **supp** For a measure F defined on $\mathcal{S}$, supp$F$ denotes the *Support* of F, which is the smallest closed subset $A \in \mathcal{S}$ such that $F(A) = 1$.

The following theorem is the major machinery we employ to study the ergodicity and mixing rate for a discrete time Markov chain.

Proposition 2.4.1: **(Nummelin and Tuominen 1982, Theorem 2.1 and 3.1)**. Let $x = (x_n)_{n\in\mathbb{N}_0}$ be a φ-irreducible aperiodic Markov chain with an n-step transition probability $P_n(dy, x)$ (the superscript $n \in \mathbb{N}_0$ is suppressed when $n = 1$), and denote the state space of x by $(\mathcal{S}, \mathcal{B}(\mathcal{S}))$, where $\mathcal{B}(\mathcal{S})$ is countably generated. Assume that there exist a measurable function $g : \mathcal{S} \to \mathbb{R}_+$, a small set $K \in \mathcal{B}(\mathcal{S})$ and constants $c_1 \in (0, 1)$ and $c_2 > 0$ such that

$$\sup_{z\in K} \int_{K^c} g(y) P(dy, z) < \infty, \tag{2.8}$$

where K^c stands for the complement of K, and that

$$\int g(y) P(dy, z) \le c_1 g(z) - c_2, \tag{2.9}$$

for any $z \in K^c$. Then x is geometrically ergodic, that is, there exists a constant $\rho \in (0, 1)$ such that

$$\int \| P_n(\cdot, z) - F \|_{TV} \, F(dz) = O(\rho^n), \qquad \text{as} \quad n \to \infty. \tag{2.10}$$

Remark From [26], the α and β mixing rates have the following relation: $2\alpha(\mathcal{U}, \mathcal{V}) \le \beta(\mathcal{U}, \mathcal{V})$. The previous result shows x is also a strong mixing process with geometric mixing rate.

Remark For the continuous time multivariate Ornstein Uhlenbeck process, Masuda (cf [52]) proved its exponential ergodicity with β-mixing rate under rather weak conditions. It turns out the technique used in the first half of the author's proof can be directly carried over to study (X_i, σ_i^2) in the BN-S model. See Lemma A.2.1 and its proof in the appendix.

We first state a supplementary result:

Lemma 2.4.2: Under the BN-S model, any compact set $A \in \mathcal{B}(\mathbb{R} \times \mathbb{R}^+)$ is a small set.

Proof: First to show the support of the joint distribution F of (X_1, σ_1^2) has a non-empty interior. Conditioning on σ_0^2 and $\{Z_s : s \in (0, h]\}$, σ_1^2 is a nonrandom function of σ_0^2 and Z_s, with X_1 being normally distributed with support on the real line. Further, the distribution of σ_0^2 is infinitely divisible and non-degenerate, so its support is unbounded (cf [63], Corollary 24.4). Therefore by unconditioning, we find F has support on $\mathbb{R} \times \mathbb{R}^+$.

It has been shown in Theorem 2.2.1 that (X_i, σ_i^2) is Weak Feller, by Lemma A.1.2 (with $\varphi = F$) we can conclude any compact set $A \in \mathcal{B}(\mathbb{R} \times \mathbb{R}^+)$ is a small set. ■

The main result in this section is the following:

Theorem 2.4.3: Let $\sigma_0^2 \overset{\mathfrak{D}}{=} \int_0^\infty e^{-s}\, \mathrm{d}Z_s$, and assume

$$E[(\sigma_0^2)^p] < \infty, \tag{2.11}$$

for some $p > 0$. Then (X_i, σ_i^2) is ergodic with geometric mixing rate.

Proof: Since (X_i, σ_i^2) is strictly stationary with $\sigma_0^2 \overset{\mathfrak{D}}{=} \int_0^\infty e^{-s}\,\mathrm{d}Z_s$, let F denote its marginal distribution, then (X_i, σ_i^2) is an F-irreducible aperiodic Markov chain. Further by Lemma 2.4.2, any compact set in $\mathcal{B}(\mathbb{R} \times \mathbb{R}^+)$ is a small set. Then by using the test function $g(x, v) = |v|^p$, the proof of Lemma A.2.1 applies and we have the geometric ergodicity of (X_i, σ_i^2). ■

Remark We use the β-mixing properties in two parts of our study: first it guarantees the existence of the asymptotic variance of the moment estimators; second, it guarantees the consistency of the kernel density estimate so we can study the limiting distribution of the minimum disparity estimate.

Remark Another well studied model which describes the joint dynamics of stock and its latent volatility is the COGARCH(1,1) model (see [44] and the reference there in). The COGARCH(1,1) process $G = (G_t)_{t \le 0}$ is defined as the solution to the SDEs:

$$\begin{aligned} dG_t &= \sigma_t\,\mathrm{d}L_t \\ d\sigma_t^2 &= (\beta - \eta\sigma_t^2)\,\mathrm{d}t \ + \varphi\sigma_t^2\,\mathrm{d}[L, L]_t^{(d)}. \end{aligned}$$

Here the G_t is the log stock price process with latent volatility σ_t. First noticing that σ_t^2 is a special case of the generalized Ornstein-Uhlenbeck processes (cf [50]) and then applying the result of Fasen (cf [28]) one concludes that σ_t^2 is exponentially β-mixing. Huag et al. (cf [35]) showed that the mixing coefficient of the increment process $G_t^{(r)} := G_t - G_{t-r} = \int_{(t-r,t]} \sigma_s dL_s$ is bounded by the mixing coefficient of

σ_t^2. This implies $G_t^{(r)}$ is α-mixing (strong mixing) with exponential mixing rate. Due to the similarity of the COGARCH(1,1) model and the BN-S model, one may conjecture that the mixing property might be proved without using the Foster-Lyapunov type criteria. We want to point out by taking our approach, we not only get the desired mixing rate, but also establish the smoothness of marginal distribution. Both components are important to study the limiting properties of MDE.

3. ESTIMATING PARAMETERS IN THE BN-S MODEL USING MINIMUM DISPARITY ESTIMATION

It is well known that traditional parametric methods such as those based on maximum likelihood are usually “automatically” optimal when the model specification is correct. But they generally suffer under model misspecification and data contamination and are poor performers from the robustness viewpoint. On the other hand, classical robust estimates such as M-estimators, which are designed be “automatically” robust for location and scale parameters, generally suffer from loss of first-order efficiency (cf Hampel et al. [32]). Although such efficiency loss is usually small, constructing a robust and efficient M-estimator for parameters other than location and scale is not always easy.

Donoho and Liu [24] studied the estimator $\hat{\theta}(P)$ based on minimizing a certain distance between a family of parametric models $\{P_\theta\}$ (indexed by θ) and the true distribution P. That is,

$$\mu(P, P_{\hat{\theta}}) = \min_{\theta} \mu(P, P_\theta),$$

where μ is a metric between probability distributions. Donoho and Liu called this kind of estimator a minimum distance estimator and they found that such estimator was automatically robust against small deviations (measured by μ) from the model

$\{P_\theta\}$. To be more specific, they showed the following:

- $\hat{\theta}(P)$ has within a factor of 2 the smallest sensitivity to small μ-perturbations among all *Fisher consistent* functionals, that is, those functionals T which satisfy $T(P_\theta) = \theta$.

- It has within a factor of 2 the best breakdown point with respect to μ-contamination among Fisher-consistent functionals.

Remark See Huber and Ronchetti [38] for more discussion on sensitivity and breakdown point.

Motivated by the pioneering work by Beran [16], Tamura and Boos [71] and Simpson [65, 66], Lindsay [49] studied in depth the efficiency and robustness of a class of minimum distance estimators, which he called the Minimum Disparity Estimators (MDE). In particular, he studied the Minimum Hellinger Distance (MHD) estimator on i.i.d. count data which follows a multinomial model. Lindsay found that the MHD method produces robust estimates while maintaining first-order (even second-order) efficiency at the true model. Another important finding is that the *influence function*, which is widely used as a measure of efficiency and robustness for the M-estimator, can be very misleading in the study of MHD. Consider the estimator (MLE, M-estimate or MDE) as a map or functional from the space of densities to the parameter space. Let this functional be denoted as T and assume it is Fisher consistent. Suppose the true distribution is $\mathbf{t}$ but what we observe is the density contaminated by amount ϵ at a fixed point ξ:

$$\mathbf{t}_\epsilon(x) := (1-\epsilon)\mathbf{t}(x) \, + \, \epsilon\chi_\xi(x)$$

where $\chi_\xi(x)$ is the indicator function for ξ. Then the quantity $\Delta T(\epsilon) := T(\mathbf{t}_\epsilon)-T(\mathbf{t})$ represents the bias caused by the contamination. Consider the Taylor approximation

$$\Delta T(\epsilon) := T(\mathbf{t}_\epsilon) - T(\mathbf{t}) \approx \, T'(\xi)\,\epsilon \tag{3.1}$$

where $T'(\xi)$ is the influence function of T defined by

$$T'(\xi) = \frac{\partial}{\partial\epsilon} T(\mathbf{t}_\epsilon)|_{\epsilon=0}. \tag{3.2}$$

Lindsay pointed out that $T'(\xi)$ played a dual role in determining the asymptotic variance of the estimate and also in controlling the magnitude of the bias. Thus if we restrict ourselves to (3.1) only, then any first-order efficient estimate which has the same $T'(\xi)$ as the MLE will be deemed as efficient but nonrobust. But from the study of MDE, Lindsay found that certain MDE's attain the optimal efficiency while retaining superior robustness compared to MLE in a location model. This led him to claim the linear approximation (3.1) is incapable of fully explaining the efficiency and robustness features of the MDE. He discovered a new class of functions, called the Residual Adjustment function (RAF) to explain this new phenomenon (more details to follow in the next section). Later, Basu and Lindsay [12] investigated the properties of MDE under continuously distributed models and showed that the

MHD estimator has bias similar to the Huber estimator while being more efficient in the location model. Basu and Sarkar [13], Basu et al. [14] and Bhandari et al. [18] extended the study to the Negative Exponential Disparity estimator (NEDE) and its generalized version (GNEDE), and they found this family of estimators achieves even better robustness against the MHD in the sense that the NEDE is also robust against inliers, that is, the outcome values predicted to be very probable by the model $\mathbf{t}$ but not expressed in the data.

In the rest of this chapter, we first summarize the findings by Basu and Lindsay in [12]. Then we will present some asymptotic results of applying the NEDE to the $\mathit{\Gamma}$-OU BN-S model.

3.1 Minimum Disparity Estimator for Continuous Models

The study by Basu and Lindsay [12] focuses on continuous models with i.i.d. data. Most of the subsequent extensions are based on this general framework. We first introduce the MDE proposed by these two authors, followed by the results which demonstrate how the MDE maintains its balance between robustness and efficiency. Finally, the consistency and asymptotic normality of the estimates are discussed.

Consider a set of i.i.d. scalar observations $\{X_1, X_2, \ldots, X_n\}$ whose CDF and density are given by $S(x)$ and $s(x)$ respectively. Assume one has a family of densities $\{m_\theta(x)\}$ indexed by an unknown parameter vector θ. Construct the kernel density estimate $f^*(x)$ by a selected known kernel $k(x;t,h)$:

$$f^*(x) = \int k(x;t,h)\,\mathrm{d}\hat{F}(t) \tag{3.3}$$

where $\hat{F}$ is the empirical distribution function. Next apply the same kernel smoothing to the model and get

$$m_\theta^*(x) = \int k(x;t,h)m_\theta(t)\,\mathrm{d}t\,. \tag{3.4}$$

Now choose a strictly convex function $G(\cdot)$ and construct a measure of "disparity" between $f^*(x)$ and $m_\theta^*(x)$:

$$\rho_G(f^*, m_\theta^*) = \int G(\delta^n(x))m_\theta^*(x)\,\mathrm{d}x \tag{3.5}$$

where

$$\delta^n(x) = (f^*(x) - m_\theta^*(x))/m_\theta^*(x) \tag{3.6}$$

is called the Pearson residual at x with the superscript n denoting its dependence on data. Then the MDE is defined to be the estimator $\hat{\theta}$ which minimizes the corresponding disparity (3.5). With different choices of G, one has several variants of the MDE, for example:

(1) Minimum Hellinger Distance (MHD):

$$HD(f^*, m_\theta^*) = \int \left[\sqrt{f^*(x)} - \sqrt{m_\theta^*(x)}\right]^2 \mathrm{d}x$$

where

$$G(\delta) = (\sqrt{\delta + 1} - 1)^2.$$

(2) Blended Weight Hellinger Distance (BWHD):

$$BWHD_\alpha(f^*, m_\theta^*) = \int \frac{(f^*(x) - m_\theta^*(x))^2}{(\alpha\sqrt{f^*(x)} - (1-\alpha)\sqrt{m_\theta^*(x)})^2} \mathrm{d}x$$

(3) Kullback-Leibler Divergence (LD):

$$LD(f^*, m_\theta^*) = \int f^*(x) \ln\left[f^*(x)/m_\theta^*(x)\right] \mathrm{d}x$$

where

$$G(\delta) = (\delta + 1)\ \ln(\delta + 1).$$

Note: in a discrete model without kernel smoothing, minimizing this divergence essentially produces the Maximum Likelihood estimator.

(4) Negative Exponential Disparity (NED):

$$NE(f^*, m_\theta^*) = \int \left(e^{-\delta^n(x)} - 1\right) m_\theta^*(x)\, \mathrm{d}x$$

where

$$G(\delta) = e^{-\delta(x)} - 1.$$

(5) Power Divergence (PD):

$$PD(f^*, m_\theta^*) = \int f^*(x)\left\{[f^*(x)/m_\theta^*(x)]^{\lambda+1} - 1\right\} \mathrm{d}x \ /\lambda(\lambda+1)$$

Remark For comparison between different disparities, see [13] and [57] for more details.

Remark Using the Pearson residual $\delta^n(x)$, the observation X_l is an outlier (or *surprising* in Basu and Lindsay) if the value of $f^*(x)/m_\theta^*(x)$ is large in its neighborhood. And it is called an inlier if the value of $f^*(x)/m_\theta^*(x)$ is close to 0.

To further study the analytic properties of the MDE, Lindsay introduced the Residual Adjustment Function (RAF) $A(\delta)$. The role of RAF is similar to the ψ-function in the M-estimator, in the sense that they both carry the efficiency and robustness information about the estimates. From the RAF, one can study the first-order, second-order (even third-order) efficiency of the estimate and investigate the trade-off between robustness and efficiency at the same time. We will discuss this feature after we introduce some definitions and concepts.

(i) **Residual Adjustment Function:** for any chosen "distance" function $G(\cdot)$

that is twice differentiable, one can define the following function

$$A(\delta) = (1+\delta)G'(\delta) - G(\delta). \tag{3.7}$$

As G is strictly convex, $A(\delta)$ is a strictly increasing function of δ. Without loss of generality, $A(\delta)$ can be centered and rescaled so that $A(0) = 0$ and $A'(0) = 1$. This centered and rescaled version of $A(\cdot)$ is called the *Residual Adjustment Function*. Further, if $A(\delta)$ is twice differentiable with $A'(\delta)$ and $A''(\delta)(1+\delta)$ which are bounded on $[-1, \infty)$, it is called *regular*.

(ii) **Transparent Kernel:** Let ∇ denote the gradient operator with respect to θ, i.e., $\nabla = \left(\partial/\partial\theta_i, \ldots, \partial/\partial\theta_p\right)^T$. If the kernel $k(x; t, h)$ satisfies the following condition:

$$C(\theta)\,\nabla \ln m_\theta(X) + D \;=\; \int \nabla \ln m_\theta^*(t) k(X; t, h)\,\mathrm{d}t$$

for all $\theta \in \Omega$ and some $p \times p$ nonsingular matrix C and p-dimensional vector D, then $k(x; t, h)$ is called a *transparent kernel* for model m_θ. A simple example is the case when m_θ is the Normal density and $k(x; t, h)$ is the Gaussian kernel (see Proposition 3.1 in [12] for more details). The advantage of using a transparent kernel is that there is no information loss when smoothing the model. However, it is generally not possible to find a transparent kernel in every model. But the simulation study conducted by Basu and Sarkar [13] showed that smoothing the data and model by the same kernel can actually increase the efficiency of MDE in some situations.

Next we will present the major findings by Basu and Lindsay. Each of the findings corresponds to a Lemma or Theorem proved by these two authors. Since these theorems are notationally heavy, the complete statements are put in the Appendix B.1 and only their implications are summarized here. Based on the study in [12], the advantages of using the MDE are as follows.

(1) Efficiency (Lemma B.1.1 and Lemma B.1.2). Basu and Lindsay showed that under some mild conditions on $A(\cdot)$, all MDE, including the LDE, have the same influence function at the model. This implies, if the kernel $k(x;t,h)$ is transparent, that the MDE achieves the same optimal variance as the MLE. Although smoothing the model by a kernel will no doubt affect the performance of the estimate, an appropriately chosen kernel will limit such efficiency loss, as demonstrated by the simulation study in [12].

(2) Robustness (Lemma B.1.2). For the MDE, one has the following approximation for the bias $\Delta T(\epsilon)$:

$$\begin{aligned} \Delta T(\epsilon) &:= T(\mathbf{t}_\epsilon) - T(\mathbf{t}) \\ &\approx T'(\xi)\epsilon + \frac{1}{2} T''(\xi)\epsilon^2 \end{aligned} \tag{3.8}$$

One notices that if the sign of $T''(\xi)$ is negative, then the bias produced by the MDE will be smaller than the one produced by the MLE. Basu and Lindsay showed that, if the model is a one parameter exponential family with θ being the location parameter and one uses the transparent kernel, then

$$T''(\xi) = A_2\, T'(\xi)\, f_2(\xi)$$

where $A_2 = A''(0)$ is called the curvature. It is not obvious under what conditions $f_2(\xi)$ and A_2 are of opposite signs. However, if one chooses a disparity which is controlled by some parameters, for example the BWHD where $A_2 = 1 - 3\alpha$, Basu and Lindsay showed that by increasing the value of α, the robustness of the estimator increases at a small cost of mean square error.

(3) Consistency and Asymptotic Normality (Lemma B.1.3). For the estimator to be consistent, one does not require the bandwidth h of the kernel density estimate $f^*(x)$ to converge to 0 as $n \to \infty$. This saves the trouble of employing different (adaptive) bandwidth selection schemes in estimating the kernel density.

3.2 Consistency and Asymptotic Normality of the NEDE in the Γ-OU BN-S model

For the Negative Exponential Disparity Estimator (NEDE), we use

$$G(\delta) = e^{-\delta} - 2$$

and find the estimate for θ by minimizing

$$\rho_{NE}(f^*, m_\theta^*) = \int \left(e^{-\delta^n(x)} - 2\right) m_\theta^*(x)\,\mathrm{d}x\,.$$

Unlike the more natural choice $G(\delta) = e^{-\delta} - 1$ which is equal to 0 when $f^* = m_\theta^*$, this specification produces a properly centered and scaled RAF $A(\delta)$ which is convenient in the study of robustness and the asymptotic normality. We see $G(\cdot)$ is a strictly convex function and bounded above by $e-2$ for $\delta \in [-1, \infty)$. As mentioned in the introduction of this chapter, the NEDE is robust against both the outliers and the inliers, and it is second-order efficient at the model in the sense of Rao (see Basu et al. [14]).

The differentiability and boundedness of $G(\cdot)$ and its derivatives make the expansion of $\rho_{NE}(f^*, m_\theta^*(x))$ easier. Compared to the general MDE, one might expect to find less stringent conditions for consistency and asymptotic normality of the estimator. But before we consider the limiting properties of the estimators, we first discuss the issue of model identifiability and the uniqueness of the estimator. These two basic concepts seem to be overlooked by many empirical studies.

Intuitively, a model g_ϑ is identifiable if different values of the parameter ϑ generate different probability distributions of the observable quantities. Since $m_{\boldsymbol{\theta}}(x)$ is the marginal density of X_i implied by the Γ-OU BN-S model, we shall approach the identifiability discussion from decomposing the original price process S_t. Recall the notions used in Chapter 1, according to [55], the dynamics of $S_t = e^{R_t}$ is given by

$$\mathrm{d}S_t = S_{t-}\big(b_t\,\mathrm{d}t + \sigma_t\,\mathrm{d}W_t + \mathrm{d}M_t\big)$$

where the appreciation rate b_t is given by the process

$$b_t = \mu + \lambda\kappa(\rho) + (\beta + \tfrac{1}{2})\sigma_t^2$$

and $M = (M_t)$ is the martingale Lévy process

$$M_t = \sum_{0<s\le t}\big(e^{\rho\Delta Z_{\lambda s}} - 1\big) - \lambda\kappa(\rho)t.$$

In a hypothetical situation where the whole trajectory of S_t is continuously observed, we can first extract the continuous parts S_t^c and the jumps part S_t^J from S_t and then identify the parameter in the following way.

- As σ_t^2 has finite activity in the Gamma OU case, one recovers σ_t^2 and the integrated volatility $\int_0^t \sigma_s^2\,\mathrm{d}s$ from the quadratic variation of S_t^c. Further notice that the marginal distribution of σ_t^2 and $\int_0^t \sigma_s^2\,\mathrm{d}s$ are uniquely determined by (ν,α) and (ν,α,λ,t) respectively. Therefore (ν,α,λ) can be identified.

- Through the jump part S_t^J of the trajectory, one can identify ρ as the cumulant transform function for the BDLP Z_t in Gamma OU is known to eqaul $\nu\rho/(\alpha - \rho)$.

- Finally, by the continuously derived b_t and σ_t^2, one can identify μ and β.

Remark A more realistic discussion of the identifiability issue is to consider that one has observations over discrete time points only. Ideally, one expects the model is still identifiable if the sampling frequency is sufficiently large. But extensive investigations over this issue in the literature seem to be missing.

In regard to the uniqueness of the estimator, we point to the following two results by Basu et al. [14] where existence and uniqueness of the NEDE are discussed. In the following exposition, let $\mathcal{G}$ denote the space of continuous densities topologized by the L_2 norm and Θ denote the parameter space. Let T_{NE} denote the Negative Exponential disparity functional, that is

$$T_{NE}(f) = \operatorname*{argmin}_{\theta \in \Theta} \rho(f, m_\theta) = \operatorname*{argmin}_{\theta \in \Theta} \int \left(e^{-\delta(x)} - 2\right) m_\theta(x)\, \mathrm{d}x \,.$$

Note: In [14], Basu et al. did not use any kernel to smooth their model density $m_\theta(x)$, so the notations in their results are un-starred.

- **Proposition** 3.2.1: **(Basu et al. 1997, Proposition 1)** Assume that

  (a) the parameter space Θ is compact;

  (b) for $\theta_1 \neq \theta_2$, $m_{\theta_1}(x) \neq m_{\theta_2}(x)$ on a set of positive Lebesgue measure;

  (c) $m_\theta(x)$ is continuous in θ for almost all x (with respect to the Lebesgue measure).

  Then

  (i) for any continuous density m, there exists a $\theta_m \in \Theta$ such that $T_{NE}(m) = \theta_m$;

  (ii) for any $\theta^* \in \Theta$, the value of $T_{NE}(m_{\theta^*})$ is unique and equal to θ^*.

- **Proposition** 3.2.2: **(Basu et al. 1997, Proposition 2)** Let $m_0(x)$ be any fixed continuous density and let $\{m_n(x)\}$ be a sequence of continuous densities. If $T_{NE}(m_0)$ is unique, then under the assumptions of Proposition 1, the functional T_{NE} is continuous at m_0 in the sense that if $m_n(x) \to m_0(x)$ in L_1, then $T_{NE}(m_n)$ converges to $T_{NE}(m_0)$ as $n \to \infty$.

Due to the similarity between these two estimation methods proposed by Basu and Lindsay [12] and Basu et al. [14], the Negative Exponential disparity to be considered may also fail to have a unique minimizer. In this paper, we will impose uniqueness assumptions on the disparity but not pursue the sufficient conditions of uniqueness.

Uniqueness Assumptions

Recall the definition of $f^*(x)$ and $m^*_{\boldsymbol{\theta}}(x)$ from (3.3) and (3.4). Let $m_{\boldsymbol{\theta}}(x)$ be the marginal density of X_1 implied by the Γ-OU BN-S model and $s^*(x)$ be the true density convolved by the kernel $k(x;t,h)$ and define $\delta^*(x) = s^*(x)/m^*_{\boldsymbol{\theta}}(x) - 1$. Assume

(**U1**) $\boldsymbol{\theta}^s$ is the unique solution to the following disparity equation in the sample space Θ.

$$\nabla\rho(s^*, m^*_\theta) \;=\; \nabla \int \big(e^{-\delta^*_s(x)} - 2\big) m^*_\theta(x)\,\mathrm{d}x \;=\; 0.$$

(**U2**) With probability approaching 1 as $n \to \infty$, $\boldsymbol{\theta}_n$ is the unique solution to the disparity equation

$$\nabla \rho(f^*, m^*_{\theta}) \;=\; \nabla \int \big(e^{-\delta^n(x)} - 2\big) m^*_{\theta}(x)\,\mathrm{d}x \;= 0$$

in a compact subset $\mathcal{K}$ of Θ which contains $\boldsymbol{\theta}^s$ and does not depend on n or data.

Remark The assumption (U1) is similar to Assumption 30 in Lindsay [49]. We point out that, when $s(x) \subseteq \{m_\theta\}$, assumption (U1) depends on the choice of the kernel. If $s(x) \nsubseteq \{m_\theta\}$, then this assumption is generally unverifiable.

For the rest of this section, fix the kernel $k(x; t, h)$ to be the Gaussian kernel and we freely use the notational simplifications,

$$\partial_i = \frac{\partial}{\partial_{\theta_i}}, \qquad \partial_{ij} = \frac{\partial^2}{\partial_{\theta_i}\partial_{\theta_j}}, \qquad \partial_i^{n_i} = \frac{\partial}{\partial \theta_i^{n_i}}.$$

Denote $\delta_s^*(x) = s^*(x)/m^*_{\boldsymbol{\theta}^s}(x) - 1$. The following result holds.

Lemma 3.2.3: Let B_a denote the 5-dimensional sphere centered at $\boldsymbol{\theta}^s$ with radius a. If $\nu > 7/2$ in the Γ-OU BN-S model, then the Taylor expansion of the Negative Exponential disparity

$$\rho(f^*, m^*_{\boldsymbol{\theta}}) \;=\; \int \big(e^{-\delta^n(x)} - 2\big) m^*_{\boldsymbol{\theta}}(x)\,\mathrm{d}x$$

with respect to $\boldsymbol{\theta}$ in the neighborhood B_a of $\boldsymbol{\theta}^s$ is given by

$$
\begin{aligned}
\rho(f^*, m^*_{\boldsymbol{\theta}}) = \rho(f^*, m^*_{\boldsymbol{\theta}^s}) &+ \sum_{i=1}^{p} \partial_i \rho(f^*, m^*_{\boldsymbol{\theta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^s}(\theta_i - \theta_i^s) \\
&+ \frac{1}{2}\sum_i \sum_j \partial_{ij}\rho(f^*, m^*_{\boldsymbol{\theta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^s}(\theta_i - \theta_i^s)(\theta_j - \theta_j^s) \\
&+ \frac{1}{6}\sum_{n_1+\dots+n_p=3} \frac{\partial^3}{\partial_1^{n_1}\dots\partial_p^{n_p}}\rho(f^*, m^*_{\boldsymbol{\theta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^s} \cdot \frac{(\theta_1-\theta_1^s)^{n_1}\dots(\theta_p-\theta_p^s)^{n_p}}{n_1!\dots n_p!} \\
&+ o_p(a^4)
\end{aligned}
\tag{3.9}
$$

where

(1)
$$
\begin{aligned}
\partial_i \rho(f^*, m^*_{\boldsymbol{\theta}})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^s} &= \int_{\mathbb{R}} \exp\big(-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1\big)\cdot\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}\cdot \partial_i m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x \\
&\quad + \int_{\mathbb{R}} \Big(\exp\big(-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1\big)-2\Big)\,\partial_i m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x
\end{aligned}
\tag{3.10}
$$

(2)
$$
\begin{aligned}
&\partial_{ij}\rho(f^*, m^*_{\boldsymbol{\theta}})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^s} \\
&= \int_{\mathbb{R}} \exp\big(-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1\big)\cdot\frac{f^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2}\cdot \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x)\cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x)\cdot m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x \\
&\quad + \int_{\mathbb{R}} \exp\big(-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1\big)\cdot\frac{f^*(x)}{m^*_{\boldsymbol{\theta}^s}(x)}\cdot \partial_{ij} m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x \\
&\quad + \int_{\mathbb{R}} \Big(\exp\big(-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1\big)-2\Big)\,\partial_{ij} m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x
\end{aligned}
\tag{3.11}
$$

(3) The third derivatives

$$
\frac{\partial^3}{\partial_1^{n_1}\dots\partial_p^{n_p}}\rho(f^*, m^*_{\boldsymbol{\theta}})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^s},
\tag{3.12}
$$

contain terms of the following forms:

- $\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{f^*(x)^3}{m^*_{\boldsymbol{\theta}^s}(x)^3} \cdot \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_k \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x$
- $\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{f^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2} \cdot \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_k \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x$
- $\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{f^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2} \cdot \partial_{ij} m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_k \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x$
- $\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{f^*(x)}{m^*_{\boldsymbol{\theta}^s}(x)} \cdot \partial_{ijk} m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x$
- $\int_{\mathbb{R}} \left(e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} - 2\right) \cdot \partial_{ijk} m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x$

Proof: See Appendix B.2 for details. ■

Remark Recall in the Γ-OU BN-S model, there are in total six parameters to be estimated: $(\lambda, \mu, \beta, \rho, \nu, \alpha)$. We estimate λ separately from the VIX data and use the estimate as the true value when estimating the remaining parameters. Thus, λ is set equal to 1 in the discussion of consistency and asymptotic normality of $(\mu, \beta, \rho, \nu, \alpha)$.

Remark The above plug-in estimator approach is valid because the density $m_{\boldsymbol{\theta}}(x)$ and its derivatives are continuous functions of λ. If one checks the steps in deriving the density $m_{\boldsymbol{\theta}}(x)$ (first part in Appendix B.2), in particular the definition (B.11) and density expression (B.17), one finds that if $\lambda \neq 1$, we need only to replace all h by λh in $m_{\boldsymbol{\theta}}(x)$ to get the completely specified density. Since h enters the $m_{\boldsymbol{\theta}}(x)$ as a constant or integration limits, by recognizing all the integrands being used in $m_{\boldsymbol{\theta}}(x)$ are continuous functions, we know that, $m_{\boldsymbol{\theta}}(x)$ and further its derivatives, are all continuous functions of λ.

To study the consistency of the NEDE, we prove the following result which considers the variance of the kernel density estimate $f^*(x)$ based on the Gaussian kernel and constructed on strong mixing data.

Lemma 3.2.4: Consider the kernel density estimate

$$f^*(x) = \frac{1}{n}\sum_{i=1}^{n} k(x; X_i, h) = \frac{1}{nh}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi}} \exp\Big[-\frac{(x - X_i)^2}{2} \Big]$$

Let $\{X_i\}$ be a strictly stationary sequence with marginal density $s(x)$. Assume that $\{X_i\}$ is strong mixing with geometric mixing rate α_m, i.e.

$$\alpha_m = O(e^{-bm})$$

for $b > 0$. Then for $0 < \Delta \ll 1$

$$\mathrm{Var}\big(f^*(x)\big) \le \frac{1}{n}s(x) + \frac{C}{n}s(x)^{\frac{2}{2+\Delta}}$$

for some constant C.

Proof: First recall a covariance estimate for strong mixing sequence given by Doukhan ([26], Section 1.2 Theorem 3):

$$\big|\mathrm{Cov}(X_i, X_j)\big| \le 8\alpha_{|i-j|}^{1/r} \, \|X_i\|_p \, \|X_j\|_q$$

for all p, q, and $r \ge 1$ with $\frac{1}{p} + \frac{1}{q} + \frac{1}{r} = 1$. Here $\|X\|_p = E[|X|^p]^{1/p}$. For a given small $\varDelta$, let $p = q = 2 + \varDelta$ so $r = \frac{2+\varDelta}{\varDelta}$. Since X_i is strictly stationary,

$$\begin{aligned}
\mathrm{Var}\big(f^*(x)\big) &= \frac{1}{n^2}\Big[\sum_{i=1}^{n} E[k(x;X_i,h)^2] + \sum\sum_{i\neq j}\mathrm{Cov}\big(k(x;X_i,h),k(x;X_j,h)\big)\Big] \\
&\leq \frac{1}{n^2}\Big[n\,E[k(x;X_1,h)^2] \\
&\qquad + \sum\sum_{i\neq j} 8\cdot\exp\big(-b|i-j|\frac{\varDelta}{2+\varDelta}\big)\cdot\big[\|k(x;X_i,h)\|_{2+\varDelta}\big]^2\Big]
\end{aligned}$$

Notice $|k(x;X_i,h)| \leq \frac{1}{\sqrt{2\pi h^2}}$ a.s., therefore

$$\begin{aligned}
\Big(\|k(x;X_i,h)\|_{2+\varDelta}\Big)^2 &= \Big(E\big[k(x;X_i,h)^{2+\varDelta}\big]\Big)^{\frac{2}{2+\varDelta}} \\
&\leq \Big((2\pi h)^{-\frac{1+\varDelta}{2}}\,E\big[k(x;X_i,h)\big]\Big)^{\frac{2}{2+\varDelta}} \\
&= \big(2\pi h\big)^{-\frac{1+\varDelta}{2+\varDelta}}\Big(E\big[k(x;X_i,h)\big]\Big)^{\frac{2}{2+\varDelta}} \\
&= \big(2\pi h\big)^{-\frac{1+\varDelta}{2+\varDelta}}\,s^*(x)^{\frac{2}{2+\varDelta}}
\end{aligned}$$

Denote $c \equiv b\frac{\varDelta}{2+\varDelta}$ and observe that

$$\sum\sum_{i\neq j}\exp\big(-c|i-j|\big) = 2\Big((n-1)e^{-c} + (n-2)e^{-2c} + \ldots + e^{-(n-1)c}\Big).$$

Let $s = (n-1)e^{-c} + (n-2)e^{-2c} + \ldots + e^{-(n-1)c}$. Then

$$\begin{aligned}
s - s\cdot e^{-c} &= (n-1)e^{-c} + (n-2)e^{-2c} + \ldots + e^{-(n-1)c} \\
&\qquad - \Big[(n-1)e^{-2c} + (n-2)e^{-3c} + \ldots + e^{-nc}\Big] \\
&= (n-1)e^{-c} - e^{-2c} - e^{-3c} - \ldots - e^{-nc} \\
&= ne^{-c} - \Big[e^{-c} + e^{-2c} + e^{-3c} + \ldots + e^{-nc}\Big].
\end{aligned}$$

Thus

$$\sum\sum_{i\neq j} \exp\big(-c|i-j|\big) = \frac{2}{1-e^{-c}}\Big(ne^{-c} - \big[e^{-c} + e^{-2c} + \ldots + e^{-nc}\big]\Big)$$
$$< \frac{2}{1-e^{-c}}\Big(ne^{-c}\Big) = 2n\frac{e^{-c}}{1-e^{-c}}.$$

This implies

$$\sum\sum_{i\neq j} 8\cdot\exp\big(-b|i-j|\frac{\Delta}{2+\Delta}\big)\cdot\big[\|k(x;X_i,h)\|_{2-\Delta}\big]^2\big]$$
$$< \frac{16ne^{-c}}{1-e^{-c}}\cdot\big(2\pi h\big)^{-\frac{1+\Delta}{2+\Delta}}\cdot s^*(x)^{\frac{2}{2+\Delta}}.$$

Let $C = (16e^{-c})/(1-e^{-c})\big(2\pi h\big)^{-\frac{1+\Delta}{2+\Delta}}$. One has

$$\begin{aligned}\mathrm{Var}\big(f^*(x)\big) &< \frac{1}{n}E[k(x;X_1,h)^2] + \frac{C}{n}s^*(x)^{\frac{2}{2+\Delta}}\\ &= \frac{1}{n}\int k(x;t,h)^2 s(t)\,\mathrm{d}t - \frac{1}{n}s^*(x) + \frac{C}{n}s^*(x)^{\frac{2}{2+\Delta}}\\ &< \frac{1}{n}s^*(x) + \frac{C}{n}s^*(x)^{\frac{2}{2+\Delta}}.\end{aligned} \tag{3.13}$$

■

Based on the previous lemma, one has the following two convergence results.

Lemma 3.2.5: If the assumptions in Lemma 3.2.4 hold, then

(1)

$$f^*(x) \xrightarrow{\mathcal{P}} s^*(x) \quad \text{as } n\to\infty$$

(2)

$$n^{1/4}\big((f^*(x))^{1/2} - (s^*(x))^{1/2}\big) \xrightarrow{\mathcal{P}} 0 \quad \text{as } n \to \infty$$

Proof: Since $\mathrm{E}[f^*(x)] = s^*(x)$, the first result is a direct consequence of the Markov inequality. Consider the second result, for any given $\epsilon > 0$, by the Chebyshev inequality:

$$\begin{aligned}
\mathbb{P}\big(n^{1/4}|f^*(x) - s^*(x)| > \epsilon\big) &= \mathbb{P}\big(|f^*(x) - s^*(x)| > n^{-1/4}\epsilon\big) \\
&\leq \frac{\mathrm{Var}(f^*(x))}{(n^{-1/4}\epsilon)^2} \\
&\to 0 \qquad \text{pointwise for each } x \text{ as} \quad n \to 0.
\end{aligned}$$

Next, expand $n^{1/4}((f^*(x))^{1/2})^2$ around $(s^*(x))^{1/2}$ for fixed x,

$$\begin{aligned}
n^{1/4}((f^*(x))^{1/2})^2 = n^{1/4}((s^*(x))^{1/2})^2 &+ 2n^{1/4}\,((s^*(x))^{1/2})\big((f^*(x))^{1/2} - (s^*(x))^{1/2}\big) \\
&+ \frac{1}{2}n^{1/4}\,\big((f^*(x))^{1/2} - (s^*(x))^{1/2}\big)^2
\end{aligned}$$

One finds that

$$\begin{aligned}
&2\,n^{1/4}\,((s^*(x))^{1/2})\big((f^*(x))^{1/2} - (s^*(x))^{1/2}\big) \\
&\quad = n^{1/4}\big(f^*(x) - s^*(x)\big) - \frac{1}{2}n^{1/4}\,\big((f^*(x))^{1/2} - (s^*(x))^{1/2}\big)^2 \\
&\quad < n^{1/4}\big(f^*(x) - s^*(x)\big) \qquad \text{a.s.}
\end{aligned}$$

From the convergence of $n^{1/4}|f^*(x) - s^*(x)|$ and the boundedness of $s(x)$, we get the desired result. ■

Let us first state the consistency result.

Theorem 3.2.6: Assume

- $\nu > 7/2$ in the model density $m_{\boldsymbol{\theta}}(x)$ described by the Γ-OU BN-S model, where $\boldsymbol{\theta} = (\mu, \beta, \rho, \nu, \alpha)$;
- $\{X_i\}$ is a strictly stationary and strong mixing scalar-valued sequence with geometric mixing rate;
- The matrix $J^{*s}(\boldsymbol{\theta}^s)$ whose ij-th element is given by (3.11) with $f^*(x)$ replaced by $s^*(x)$ is a positive definite matrix.

Then, the NEDE $\boldsymbol{\theta}_n \xrightarrow{\mathcal{P}} \boldsymbol{\theta}^s$ as $n \to \infty$.

Proof: using the similar arguments in Lehmann and Casella [47], one considers the behavior of $\rho(f^*, m^*_{\boldsymbol{\theta}})$ on the sphere B_a centered at $\boldsymbol{\theta}^s$ with radius a. We will show that for any sufficiently small a,

$$\min_{\boldsymbol{\theta} \in B_a} \left(\rho(f^*, m^*_{\boldsymbol{\theta}}) - \rho(s^*, m^*_{\boldsymbol{\theta}^s})\right) > \frac{1}{2} a^T J^{*s}(\boldsymbol{\theta}^s) a \tag{3.14}$$

with probability converging to 1. This implies that for any $a > 0$, as $n \to \infty$, the minimum disparity equation for $\rho(f^*, m^*_\theta)$ attains its local minimum in B_a at θ_n with probability tending to 1.

By Appendix B.2, all the coefficients of the Taylor expansion listed in Lemma 3.2.3 are absolutely integrable, independent of $f^*(x)$. This means we can apply the

Dominated convergence theorem (DCT) to each coefficient. For example, consider the following term in (3.11):

$$\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{f^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2} \cdot \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x \, .$$

Observe that

$$\left| \exp\left(- \frac{f^*}{m^*_{\boldsymbol{\theta}^s}} + 1 \right) \cdot \frac{f^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2} \right| \leq 2$$

is bounded by 2 independently of $f^*(x)$ and

$$\int_{\mathbb{R}} \left| \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x) \right| \cdot m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x \; < \infty$$

due to Proposition B.2.6 and Lemma B.2.12. Then by the Dominated convergence theorem, we have

$$\begin{aligned} &\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{f^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2} \cdot \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\theta^s}(x) \cdot m^*_{\theta^s}(x)\, \mathrm{d}x \\ \xrightarrow{\mathcal{P}} &\int_{\mathbb{R}} e^{-\frac{s^*}{m^*_{\theta^s}}+1} \cdot \frac{s^*(x)^2}{m^*_{\boldsymbol{\theta}^s}(x)^2} \cdot \partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x \end{aligned}$$

as $n \to \infty$. Similarly we can show the convergence for the rest of the coefficients. Therefore, for terms in (3.10),

$$\begin{aligned} \int_{\mathbb{R}} e^{-\frac{s^*}{m^*_{\boldsymbol{\theta}^s}}+1} \cdot \frac{s^*}{m^*_{\boldsymbol{\theta}^s}} \cdot \partial_i m^*_{\theta^s}(x)\, \mathrm{d}x \; + \int_{\mathbb{R}} (e^{-\frac{s^*}{m_{\theta^s}}+1} - 2)\, \partial_i m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x \\ = \int_{\mathbb{R}} A(\delta^*_s(x))\, \partial_i m^*_{\boldsymbol{\theta}^s}(x)\, \mathrm{d}x \;\; = \; 0 \end{aligned}$$

by the definition of $\boldsymbol{\theta}^s$, so the linear terms in (3.9) are of order a^3 for large n. On the other hand, terms in (3.11) and (3.12) all converge to finite limits. This implies that the cubic terms in (3.9) are of order a^3 for large n. Finally, given that $\boldsymbol{J}^{*s}(\boldsymbol{\theta}^s)$ is positive definite,

$$\min_{\boldsymbol{\theta}\in B_a} \left(\rho(f^*, m^*_{\boldsymbol{\theta}}) - \rho(s^*, m^*_{\boldsymbol{\theta}^s})\right) > \frac{1}{2}a^T J^{*s}(\boldsymbol{\theta}^s)a.$$

Therefore, for any small value a,

$$\rho(f^*, m^*_{\boldsymbol{\theta}}) > \rho(s^*, m^*_{\boldsymbol{\theta}^s})$$

for all $\boldsymbol{\theta}$ on the surface of $\boldsymbol{B_a}$ for sufficiently large n. Since $\boldsymbol{\theta}_n$ solves the minimum disparity equation, i.e., minimizes $\rho(f^*, m^*_{\boldsymbol{\theta}})$, this means with probability approaching 1, the local minimizer of $\rho(f^*, m^*_{\boldsymbol{\theta}})$ is in the interior of $\boldsymbol{B_a}$. The consistency of $\boldsymbol{\theta}_n$ is proved. ■

Next we discuss the asymptotic normality of the NEDE. In order to prove the central limit theorem for NEDE, we first derive the asymptotic distribution of

$$\sqrt{n}\int_{\mathbb{R}} \left[A(f^*/m^*_{\boldsymbol{\theta}^s} - 1) - A(s^*/m^*_{\boldsymbol{\theta}^s} - 1)\right]\nabla m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x$$

where $A(\delta) = 2-(2+\delta)e^{-\delta}$. Since the data $\{X_i\}$ is a stationary β-mixing sequence, the following result by Ibragimov and Linnik [39] is useful.

Lemma 3.2.7: **(Ibragimov and Linnik 1971, Theorem 18.5.3)** Let the mean zero stationary sequence X_j satisfy the strong mixing condition with mixing coefficient $\alpha(n)$, and let $E|X_j|^{2+\delta} < \infty$ for some $\delta > 0$. If

$$\sum_{n=1}^{\infty} \alpha(n)^{\delta/(2+\delta)} < \infty,$$

then

$$\sigma^2 \;=\; E(X_0^2) + 2\sum_{j=1}^{\infty} E(X_0\, X_j) \;<\; \infty,$$

and if $\sigma \neq 0$, then

$$\lim_{n\to\infty} \mathbf{P}\Big\{\sigma^{-1}\,\frac{1}{\sqrt{n}}\,\sum_{j=1}^{n} X_j \;<\; z\Big\} = \Phi(z),$$

where $\Phi(z)$ is the standard Normal CDF.

Lemma 3.2.8: Let $\{X_i\}$ satisfy the conditions in Lemma 3.2.4. Assume that the RAF $A(\delta)$ is regular and condition

$$\int s^*(x)^{\frac{1}{2+\Delta}}\,|\nabla\, \ln m^*_{\theta^s}(x)|\, \mathrm{d}x \;<\infty. \tag{3.15}$$

holds for $0 < \Delta \ll 1$. Further assume that

$$E\Big|\int k(x,X,h)\, A'(\delta^*_s(x))\, \nabla \ln m^*_{\theta^s}(x)\, \mathrm{d}x\,\Big|^{\otimes(2+c)} < \infty \tag{3.16}$$

holds for some $c > 0$. Then

$$n^{1/2}\int\Big[A(\delta_n) \;-\; A(\delta^*_s)\Big]\nabla m^*_{\theta^s}(x)\, \mathrm{d}x \;\to MVN(0,V) \tag{3.17}$$

where V is given by

$$V \;=\; E(V_0^{\otimes 2}) + 2\sum_{j=1}^{\infty} E(V_0\, V_j^T) \;<\; \infty,$$

with T denoting the transpose of a vector and

$$V_j \;=\; \int k(x, X_j, h)\; A'(\delta_s^*(x))\; \nabla\; \ln m_{\theta^s}^*(x)\, \mathrm{d}x\,.$$

Remark Given the variance upper bound in Lemma 3.2.5, we proceed through the proof by following the approach in Basu and Lindsay ([12], Section 6).

Proof: Define the Hellinger residual Δ_n and Δ_s^* as

$$\Delta_n \;=\; \frac{(f^*(x))^{1/2}}{m_{\boldsymbol{\theta}^s}^{*1/2}} - 1 \quad \text{and} \quad \Delta_s^* \;=\; \frac{(s^*(x))^{1/2}}{m_{\boldsymbol{\theta}^s}^{*1/2}} - 1.$$

Let $Y_n(x) \;=\; n^{1/2}(\Delta_n(x) - \Delta_s^*(x))^2$. Since for $a, b \;\geq\; 0$, $(\sqrt{a} - \sqrt{b})^2 \;\leq\; |a - b|$, therefore, for $k \in [0, 2]$,

$$\begin{aligned} E[Y_n^k] &= E\Big[n^{k/2}\,\big(\frac{(f^*(x))^{1/2} - (s^*(x))^{1/2}}{m_{\boldsymbol{\theta}^s}^{*1/2}}\big)^{2k}\Big] \\ &\leq \frac{n^{k/2}}{m_{\boldsymbol{\theta}^s}^*(x)^k}\; E\Big[\big|f^*(x) - s^*(x)\big|^k\Big] \\ &\leq \frac{n^{k/2}}{m_{\boldsymbol{\theta}^s}^*(x)^k}\; E\Big[\big|f^*(x) - s^*(x)\big|^2\Big]^{k/2} \qquad (\triangle) \end{aligned}$$

By (3.13),

$$\begin{aligned}\triangle &\leq \frac{n^{k/2}}{m^*_{\boldsymbol{\theta}^s}(x)^k}\left(\frac{C}{n}\left(s^*(x)+s^*(x)^{\frac{1}{2+\Delta}}\right)\right)^{k/2}\\ &= \frac{C^{k/2}}{m^*_{\boldsymbol{\theta}^s}(x)^k}\left(s^*(x)+s^*(x)^{\frac{1}{2+\Delta}}\right)^{k/2} < \infty\end{aligned}$$

The third line holds due to Lyapunov's inequality. From Lemma 3.2.5 we know $Y_n \to 0$ in probability and we just show $\sup_n E[Y_n^k]$ is bounded for $k \in [0,2)$, then

$$\lim_{n\to\infty} E[Y_n^p] = 0 \qquad \text{for} \qquad p \in [0,2). \tag{3.18}$$

Next introduce some notations:

- $a_n(x) = A(\delta_n(x)) - A(\delta_s^*(x))$ and $b_n(x) = (\delta_n(x) - \delta_s^*(x))A'(\delta_s^*(x))$.
- $\gamma_n = \int n^{1/2}(a_n(x) - b_n(x))\nabla m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x$ and $\tau_n = n^{1/2}|a_n(x) - b_n(x)|$.

By using the analytic property of a regular RAF $A(\delta)$, Lindsay (1994, Lemma 25) proved

$$E[\tau_n(x)] \leq BE[Y_n(x)] \qquad \text{for} \qquad B > 0.$$

From (3.18), one can conclude $E[\tau_n(x)] \to 0$. Now

$$\begin{aligned}E|\gamma_n| &\leq \int E(\tau_n(x))|\nabla m^*_{\boldsymbol{\theta}^s}(x)|\,\mathrm{d}x\\ &\leq \int \left(\left(s^*(x)+s^*(x)^{\frac{1}{2+\Delta}}\right)\right)^{1/2}\cdot\frac{C^{1/2}}{m^*_{\boldsymbol{\theta}^s}(x)}|\nabla m^*_{\boldsymbol{\theta}^s}(x)|\,\mathrm{d}x\\ &= C^{1/2}\int \left(\left(s^*(x)+s^*(x)^{\frac{1}{2+\Delta}}\right)\right)^{1/2}\cdot|\nabla \ln m^*_{\boldsymbol{\theta}^s}(x)|\,\mathrm{d}x\end{aligned} \tag{3.19}$$

Since we have shown in Lemma B.2.12 that

$$|\nabla \ln m^*_{\boldsymbol{\theta}}(x)| \;\leq\; M(h,\boldsymbol{\theta})\,(1+|x|^l)$$

for some positive l and $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}^s$. Then the integral (3.19) is finite when (3.15) holds. Notice that (3.19) is independent of n, therefore we can use the Dominated convergence theorem to conclude

$$n^{1/2}\int \Big|A(\delta_n(x)) - A(\delta^*_s(x)) - (\delta_n(x)-\delta^*_s(x))A'(\delta^*_s(x))\Big|\nabla m^*_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \;\to\; 0$$

as $n\to\infty$. This means we can find the asymptotic distribution of (3.17) by studying the limiting distribution of

$$\begin{aligned}
&n^{1/2}\int \big(\delta_n(x)-\delta^*_s(x)\big)A'(\delta^*_s(x))\nabla m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x\\
&= n^{1/2}\Big[\frac{1}{n}\int\sum_1^n k(x;X_i,h)A'(\delta^*_s(x))\nabla\ln m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x\\
&\qquad -\int\int k(x;t,h)s(t)\,\mathrm{d}t\;\;A'(\delta^*_s(x))\nabla\ln m^*_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x\,\Big]
\end{aligned}$$

One finds the above expression is in fact the root-n normalized sum of n mean zero strongly mixing random vectors. By using Lemma 3.2.7 and the Cramér-Wold device, it is asymptotically normal with mean 0 and variance-covariance matrix given by V. ■

Theorem 3.2.9: Assume the conditions in Theorem 3.2.6 and Lemma 3.2.8 hold.

Then the NEDE $\boldsymbol{\theta}_n$ satisfies

$$\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}^s) \xrightarrow{\mathcal{P}} MVN(0, V_{NE})$$

where

$$V_{NE} = J^*(\boldsymbol{\theta}^s)^{-1} V J^*(\boldsymbol{\theta}^s)^{-1}.$$

Proof: The proof is carried out by first performing a second order Taylor expansion to

$$\nabla \int \left(e^{-f^*(x)/m_{\boldsymbol{\theta}}(x)+1} - 2 \right) \mathrm{d}x$$

with respect to $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}^s$. Recall the consistency of $\boldsymbol{\theta}_n$ proved in Theorem 3.2.6 and the asymptotic result in Lemma 3.2.8, then the arguments in Lehmann and Casella ([47], Theorem 5.1 (b), p 464) apply. ■.

Remark From Lemma B.1.1, the asymptotic variance-covariance matrix of $\boldsymbol{\theta}_n$ is independent of $G(\delta) = \exp(-\delta) - 2$ when $s(x) \subseteq \{m_{\boldsymbol{\theta}}(x)\}$. Since $m_{\boldsymbol{\theta}}(x)$ in the Γ-OU BN-S model can be thought of a continuous mixture of Gaussian distributions with different means and variances, we conjecture that the efficiency loss due to the use of Gaussian kernel should be limited.

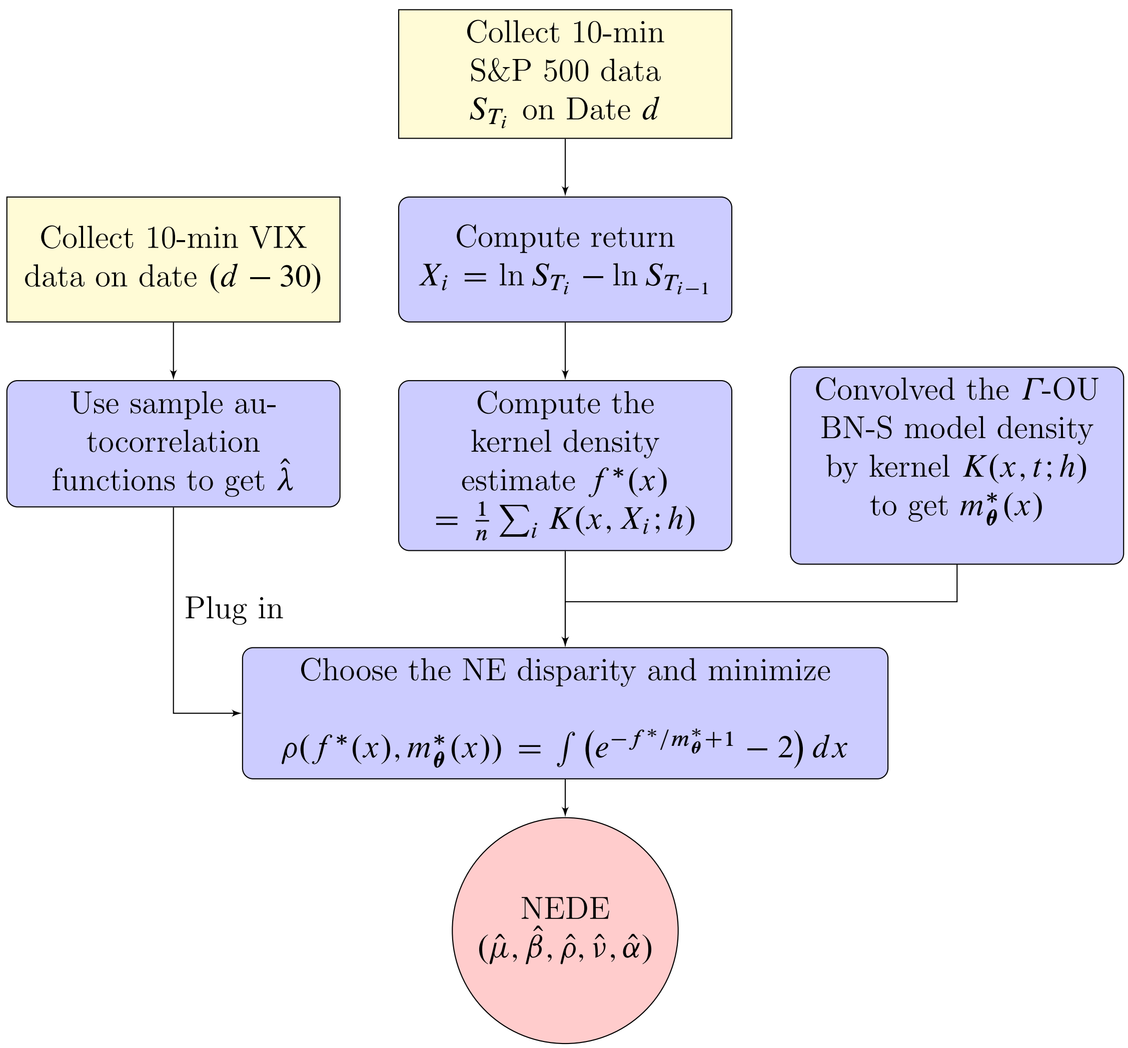
At the end of this chapter, we use a diagram (in next page) to illustrate how to implement the NEDE with the Γ-OU BN-S model. There are several details worth to mention first.

(1) Since the VIX is the expected future volatility, the VIX values which correspond

to the current values of S_t are those ones 30 days (or 22 business days) ago.

(2) The data sampling frequency for the VIX data should be no lower than the frequency for the S&P 500 data so to make sure the plug in estimator $\hat{\lambda}$ converges at the same speed as the NEDE.

(3) Although there is no explicit form for $m_{\boldsymbol{\theta}}(x)$, one can jointly simulate (X_i, σ_i^2) to compute $m_{\boldsymbol{\theta}}(x)$ numerically. Since the BDLP process Z_t is a compound Poisson process in the Γ-OU case, we can simulate (X_i, σ_i^2) very efficietly.



4. ESTIMATING THE PARAMETERS IN THE BN-S MODEL USING MOMENT BASED METHODS

From the study in Chapter 2, we have shown that (X_i, σ_i^2) is strictly stationary and β-mixing with geometric mixing rate. This enables us to use Birkhoff's ergodic theorem to study the limiting behavior of the moment estimators. However, as one can not observe σ_t^2 in the empirical study, the conventional method of moments can not be used unless other quantities known to be linearly dependent on σ_t^2 are available. But if one is only interested in estimating the parameters in the volatility components, then the estimators based on the realized multipower variations can be used. In this section, we will discuss how to construct the MOM estimators and study their asymptotic properties if (X_i, σ_i^2) are both observed.

Recall again we observe the processes R_t and σ_t^2 in a finite time horizon $[0, T]$ over $(n+1)$ equi-spaced time points $T_i = i\frac{T}{n}$ for $i = 0, 1, \ldots, n$. The bivariate series (X_i, σ_i^2) where $X_i = R_{T_i} - R_{T_i-1}$ has its dynamics defined by (2.1):

$$\begin{cases} X_i \;=\; \mu h + \beta \int_{(i-1)h}^{ih} \sigma_s^2 \,\mathrm{d}s \;+\; \int_{(i-1)h}^{ih} \sigma_s \,\mathrm{d}W_s \;+\; \rho \int_{(i-1)h}^{ih} \mathrm{d}Z_{\lambda s}\,, \\ \sigma_i^2 \;=\; e^{-\lambda h}\sigma_{i-1}^2 \;+\; \int_{(i-1)h}^{ih} e^{-\lambda(ih-s)}\,\mathrm{d}Z_{\lambda s}\,. \end{cases}$$

If $\sigma_0^2 \overset{\mathcal{D}}{=} \int_0^{-\infty} e^{-s}\,\mathrm{d}Z_s$, then the bivariate series is strictly stationary. There are two features in the BN-S model that we should keep in mind when designing the estimation scheme:

- Let $\boldsymbol{\eta}$ denote the parameters in the distribution of Z_t. In the model specification above, the marginal distribution of σ_i^2 is independent of λ while its autocorrelation function only depends on λ. So $\boldsymbol{\eta}$ and the mean reverting parameter λ can be estimated solely from $\{\sigma_i^2\}$.

- The characteristic function of X_i is known explicitly for the Γ-OU and IG-OU cases, but it is very complicated and it is impractical to derive moments of X_i of order higher than 2.

Denoting the discretely observed squared volatility σ_i^2 by V_i. We propose to estimate $\boldsymbol{\theta} = (\boldsymbol{\eta}, \lambda, \beta, \mu, \rho)$ by the following algorithm:

Step 1: Estimate λ Recall that, as discussed in Section 1.2, we know V_i is a strictly stationary series with finite mean and variance. Its autocorrelation function is given by

$$\mathrm{Corr}\big(V_i, V_j\big) = e^{-\lambda|i-j|}.$$

Fix $i = 0$ and let j range from 0 to $d < n$. Define the lagged-j sample autocovariance functions and sample autocorrelation functions by

$$\hat{\varphi}_n(j) = \frac{1}{n}\sum_{k=1}^{n-j}(V_{k+j} - \bar{V})(V_k - \bar{V}) \qquad \text{and} \qquad \hat{r}_n(j) = \frac{\hat{\varphi}_n(j)}{\hat{\varphi}_n(0)}.$$

Here $\bar{V} = \frac{1}{n}\sum_{k=1}^{n} V_k$. One finds $\hat{\varphi}_n(0) = \frac{1}{n}\sum_{k=1}^{n}(V_k - \bar{V})^2$ is the MOM estimator for $\mathrm{Var}(\sigma_0^2)$. Denote $(\hat{\varphi}_n(1), \ldots, \hat{\varphi}_n(d))$ by $\hat{\boldsymbol{\varphi}}$. If there exists a sequence of $\hat{\lambda}_n$ such that

$$\hat{\lambda}_n = \underset{\lambda>0}{\operatorname{argmin}} \sum_{j=1}^{d} \left(\hat{r}_n(j) - e^{-\lambda j}\right)^2,$$

then, according to Spiliopoulos [67],

$$\hat{\lambda}_n \to \lambda \qquad \text{a.s. as} \quad n \to \infty. \tag{4.1}$$

In what follows, we suppress the subscript n in the estimators.

Step 2: Estimate $\boldsymbol{\eta}$ Here we avoid a general discussion but focus on the Γ-OU and the IG-OU BN-S models, where $\boldsymbol{\eta}$ equals (ν, α) and (δ, γ) respectively. Since the marginal distribution of V_i is independent of λ, $\boldsymbol{\eta}$ can be directly estimated by the first two absolute moments of $\{V_i\}$ without plugging in $\hat{\lambda}$. Further, the marginal distribution for V_i in the Γ-OU (IG-OU) BN-S model is simply the Gamma(ν, α) (IG(δ, γ)) distribution, moments of which up to fourth order can be computed efficiently by using the characteristic function. For a Gamma(ν, α) random variable G:

$$E[G] = \tfrac{\nu}{\alpha}, \qquad E[G^2] = \tfrac{\nu(\nu+1)}{\alpha^2}$$
$$E[G^3] = \tfrac{\nu(\nu^2+3\nu+2)}{\alpha^3}, \qquad E[G^4] = \tfrac{\nu(\nu^3+6\nu^2+11\nu+6)}{\alpha^4}$$

and for a IG(δ, γ) random variable L:

$$E[L] = \tfrac{\delta}{\gamma}, \qquad E[L^2] = \tfrac{\delta(\delta\gamma+1)}{\gamma^3}$$
$$E[L^3] = \tfrac{\delta(\delta^2\gamma^2+3\delta\gamma+3)}{\gamma^5}, \qquad E[L^4] = \tfrac{\delta(\delta^3\gamma^3+6\delta^2\gamma^2+15\delta\gamma+15)}{\gamma^7}$$

Let

$$\overline{V^2} = \frac{1}{n}\sum_{k=1}^n V_k^2$$

In the Γ-OU BN-S model, we solve

$$\begin{cases} \bar{V} = \dfrac{\nu}{\alpha} \\ \\ \overline{V^2} = \dfrac{\nu(\nu+1)}{\alpha^2} \end{cases} \quad \text{and get} \quad \begin{cases} \hat{\alpha} = \frac{\bar{V}}{\overline{V^2}-(\bar{V})^2} \\ \\ \hat{\nu} = \frac{(\bar{V})^2}{\overline{V^2}-(\bar{V})^2} \end{cases} \tag{4.2}$$

In the IG-OU BN-S model, we solve

$$\begin{cases} \bar{V} = \dfrac{\delta}{\gamma} \\ \\ \overline{V^2} = \dfrac{\delta(\delta\gamma+1)}{\gamma^3} \end{cases} \quad \text{and get} \quad \begin{cases} \hat{\delta} = \sqrt{\frac{(\bar{V})^3}{\overline{V^2}-(\bar{V})^2}} \\ \\ \hat{\gamma} = \sqrt{\frac{\bar{V}}{\overline{V^2}-(\bar{V})^2}} \end{cases} \tag{4.3}$$

Step 3: Estimate (β, μ, ρ) We need the the covariance of (X_1, σ_1^2) and moments of X_1 to find these estimators. Letting $h = 1$ in equation (C.5), one has

$$\mathrm{Cov}(X_1, \sigma_1^2) = \left(\frac{\beta}{\lambda} + 2\rho\right)\left(1 - e^{-\lambda h}\right)\mathrm{Var}(\sigma_0^2).$$

Under the Γ-OU BN-S Model, the mean and variance of X_1 are given by (C.2):

$$E[X_1] = h\mu + \frac{h\nu(\beta+\lambda\rho)}{\alpha}$$
$$\mathrm{Var}(X_1) = \frac{\nu}{\alpha^2\lambda^2}\Big((2\beta^2 + 4\beta\lambda\rho)(e^{-\lambda h} + (\lambda h - 1)) + h\lambda^2(\alpha + 2\lambda\rho^2)\Big).$$

From $\mathrm{Cov}(X_1, \sigma_1^2)$ and $\mathrm{Var}(X_1)$ the estimator of (β, ρ) can be derived by solving a system of equations with the restriction $\rho < 0$:

$$\begin{aligned}\hat{\beta} =& \frac{1}{\hat{\nu}h\,(e^{\hat{\lambda}h}-1)^2\,\widehat{\mathrm{Var}}(\sigma_0^2)}\Big[\sqrt{2\Lambda_1} \\ &+ \hat{\nu}\,(1-e^{\hat{\lambda}h})\left(2+e^{\hat{\lambda}h}(\hat{\lambda}h-2)\right)\widehat{\mathrm{Var}}(\sigma_0^2)\,\widehat{\mathrm{Cov}}(X_1,\sigma_1^2)\Big] \qquad (4.4)\\ \hat{\rho} =& \frac{1}{2\,\hat{\nu}\hat{\lambda}h\,(e^{\hat{\lambda}h}-1)^2\,\widehat{\mathrm{Var}}(\sigma_0^2)}\Big[-\sqrt{2\Lambda_1} \\ &+ 2\,\hat{\nu}\,(-1+e^{\hat{\lambda}h})\left(1+e^{\hat{\lambda}h}(\hat{\lambda}h-1)\right)\widehat{\mathrm{Var}}(\sigma_0^2)\,\widehat{\mathrm{Cov}}(X_1,\sigma_1^2)\Big] \qquad (4.5)\end{aligned}$$

where

$$\begin{aligned}\Lambda_1 =& \hat{\nu}\left(e^{\hat{\lambda}h}-1\right)^3\widehat{\mathrm{Var}}(\sigma_0^2)\left[2\,\hat{\nu}\left(1+e^{\hat{\lambda}h}(\hat{\lambda}h-1)\right)\widehat{\mathrm{Cov}}^2(X_1,\sigma_1^2)\right. \\ &\left.+\ \hat{\alpha}\hat{\lambda}h\,(e^{\hat{\lambda}h}-1)\left(\hat{\nu}h-\hat{\alpha}\widehat{\mathrm{Var}}(X_1)\right)\widehat{\mathrm{Var}}^2(\sigma_0^2)\right]\end{aligned}$$

and

$$\begin{aligned}\widehat{\mathrm{Var}}(\sigma_0^2) &= \overline{V^2} - (\bar{V})^2 \\ \overline{XV} \triangleq \widehat{\mathrm{Cov}}(X_1,\sigma_0^2) &= \frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X})(V_k-\bar{V}) \\ \hat{\sigma}_X^2 \triangleq \widehat{\mathrm{Var}}(X_1) &= \frac{1}{n}\sum_{k=1}^{n}(X_k-\bar{X})^2 \\ \text{with} \qquad \bar{X} &= \frac{1}{n}\sum_{k=1}^{n}X_k = \frac{1}{n}(R_n-R_0)\end{aligned}$$

Then $\hat{\mu}$ can be obtained from $E[X_1]$:

$$\hat{\mu} = \frac{1}{h}\bar{X} - \frac{\hat{\nu}}{\hat{\alpha}}(\hat{\beta}+\hat{\lambda}\hat{\rho}) \qquad (4.6)$$

Under the IG-OU BN-S Model, the mean and variance of X_1 are given by (C.3)

$$
\begin{aligned}
E[X_1] &= \frac{h(\gamma\mu + \beta\delta + \delta\lambda\rho)}{\gamma} \\
\mathrm{Var}[X_1] &= \frac{\delta}{\gamma^3\lambda^2}\Big((2\beta^2 + 4\beta\lambda\rho)(e^{-\lambda h} + (\lambda h - 1)) + h\lambda^2\big(\gamma^2 + 2\lambda\rho^2\big)\Big)
\end{aligned}
$$

solving the corresponding system of equations with the restriction $\rho < 0$:

$$
\begin{aligned}
\hat{\beta} = \; & \frac{1}{\hat{\delta} h\,(e^{\hat{\lambda} h} - 1)^2\,\widehat{\mathrm{Var}}(\sigma_0^2)}\Big[\sqrt{2\Lambda_2} \\
& + \hat{\delta}\,(1 - e^{\hat{\lambda} h})\,\big(2 + e^{\hat{\lambda} h}(\hat{\lambda} h - 2)\big)\,\widehat{\mathrm{Var}}(\sigma_0^2)\,\widehat{\mathrm{Cov}}(X_1, \sigma_1^2)\Big] \qquad (4.7) \\
\hat{\rho} = \; & \frac{1}{2\,\hat{\delta}\hat{\lambda} h\,(e^{\hat{\lambda} h} - 1)^2\,\widehat{\mathrm{Var}}(\sigma_0^2)}\Big[-\sqrt{2\Lambda_2} \\
& + 2\,\hat{\delta}\,(-1 + e^{\hat{\lambda} h})\,\big(1 + e^{\hat{\lambda} h}(\hat{\lambda} h - 1)\big)\,\widehat{\mathrm{Var}}(\sigma_0^2)\,\widehat{\mathrm{Cov}}(X_1, \sigma_1^2)\Big] \qquad (4.8)
\end{aligned}
$$

where

$$
\begin{aligned}
\Lambda_2 = \; & \hat{\delta}\,\big(e^{\hat{\lambda} h} - 1\big)^3\,\widehat{\mathrm{Var}}(\sigma_0^2)\,\Big[2\,\hat{\delta}\,\big(1 + e^{\hat{\lambda} h}(\hat{\lambda} h - 1)\big)\,\widehat{\mathrm{Cov}}^2(X_1, \sigma_1^2) \\
& + \hat{\gamma}^2 h\,(e^{\hat{\lambda} h} - 1)\,\big(\hat{\delta} h - \hat{\gamma}\,\widehat{\mathrm{Var}}(X_1)\big)\Big]
\end{aligned}
$$

And $\hat{\mu}$ can be obtained from $E[X_1]$:

$$
\hat{\mu} = \frac{1}{h}\bar{X} - \frac{\hat{\delta}}{\hat{\gamma}}(\hat{\beta} + \hat{\lambda}\hat{\rho}) \qquad (4.9)
$$

At last, we discuss the consistency and asymptotic normality of the MOM estimators. Our first result considers the strong consistency of the sample moments and the MOM estimators.

Theorem 4.0.10: For the (X_i, σ_i^2) considered in the Γ-OU BN-S model and the IG-OU BN-S model,

(i) The sample moments are strongly consistent, i.e., as $n \to \infty$

- $\bar{V} \xrightarrow{a.s.} E[\sigma_0^2]$ and $\overline{V^2} \xrightarrow{a.s.} E[(\sigma_0^2)^2]$;
- $\hat{\varphi}_n(j) \xrightarrow{a.s.} \varphi(j)$ for $j = 1, \ldots, d$;
- $\overline{XV} \xrightarrow{a.s.} \mathrm{Cov}(X_1, \sigma_0^2)$;
- $\bar{X} \xrightarrow{a.s.} E[X_1]$ and $\hat{\sigma}_X^2 \xrightarrow{a.s.} \mathrm{Var}(X_1)$.

(ii) The MOM estimator $\hat{\boldsymbol{\theta}}_M \triangleq (\hat{\boldsymbol{\eta}}, \hat{\beta}, \hat{\mu}, \hat{\rho}, \hat{\lambda})^T$ is strongly consistent, that is

$$\hat{\boldsymbol{\theta}}_M \xrightarrow{a.s.} (\boldsymbol{\eta}, \beta, \mu, \rho, \lambda)^T \qquad \text{as} \qquad n \to \infty.$$

Proof: The first result is a direct application of the Birkhoff's ergodic theorem. For the second result, recall that $\hat{\lambda}$ is strongly consistent for λ due to Spiliopoulos [67]. The strong consistency of $(\hat{\boldsymbol{\eta}}, \hat{\beta}, \hat{\mu}, \hat{\rho})$ under the Γ-OU (or IG-OU) BN-S model comes from the fact that, if we replace the sample moments in equation (4.2), (4.4) and (4.6) (or (4.3), (4.7) and (4.9)) by the corresponding population moments, then the parameters $(\boldsymbol{\eta}, \beta, \mu, \rho)$ are continuous functions of the population moments. Therefore, by the continuous mapping theorem, $(\hat{\boldsymbol{\eta}}, \hat{\beta}, \hat{\mu}, \hat{\rho})$ are strongly consistent.

■

Next we show the asymptotic normality of the sample moments.

Theorem 4.0.11: For the (X_i, σ_i^2) considered in the $\mathit{\Gamma}$-OU BN-S model and the IG-OU BN-S model, the sample moments $\widehat{\Upsilon} = (\bar{V}, \overline{V^2}, \hat{\boldsymbol{\varphi}}, \overline{XV}, \bar{X}, \hat{\sigma}_X^2)$ are asymptotically normal:

$$\sqrt{n}\big(\widehat{\Upsilon} - \Upsilon\big) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \boldsymbol{\Sigma}) \tag{4.10}$$

where

$$\Upsilon = \big(E[\sigma_0^2],\ E[(\sigma_0^2)^2],\ \boldsymbol{\varphi},\ \mathrm{Cov}(X_1, \sigma_0^2), E[X_1], \mathrm{Var}(X_1)\big)$$

and $\boldsymbol{\Sigma}$ is the variance-covariance matrix given by

$$\boldsymbol{\Sigma} = E\big[\mathbf{U}_0^{\otimes 2}\big] + \sum_{k=1}^{\infty} E\big[\mathbf{U}_0\mathbf{U}_k^T\big]$$

with the $d+5$ dimension vector $\mathbf{U}_i$ defined as

$$\mathbf{U}_i = \Big(V_i,\ V_i^2,\ (V_{i+1} - E[V_1])(V_i - E[V_1]), \ldots, (V_{i+d} - E[V_1])(V_i - E[V_1]),$$
$$(X_i - E[X_1])(V_i - E[V_1]),\ X_i,\ (X_i - E[X_1])^2\Big)^T$$

Proof: Since in the $\mathit{\Gamma}$-OU BN-S model and the IG-OU BN-S model, all moments of X_i and V_i are finite, the proof of Proposition 2 in Haug et al. [35] can be directly carried over to our study with $\mathbf{Y}_i$ in their proof replaced by $\mathbf{U}_i$, and then (4.10) follows. ■

Let $\mathcal{H}$ denote the mapping from Υ to $\boldsymbol{\theta} = (\boldsymbol{\eta}, \lambda, \beta, \mu, \rho)$ defined by equations (4.2), (4.1), (4.4) and (4.6) in the $\mathit{\Gamma}$-OU BN-S model (or (4.3), (4.1), (4.7) and (4.9) in the IG-OU BN-S model) with the sample moments replaced by the population moments. We have the following asymptotic result of the MOM estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{M}}$.

Theorem 4.0.12: The MOM estimator $\hat{\boldsymbol{\theta}}_{\boldsymbol{M}} = (\hat{\boldsymbol{\eta}}, \hat{\lambda}, \hat{\beta}, \hat{\mu}, \hat{\rho})$ is asymptotically normal:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{\boldsymbol{M}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} MVN(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{M}}) \tag{4.11}$$

as $n \to \infty$. Where $\boldsymbol{\Sigma}_{\boldsymbol{M}}$ is given by

$$\boldsymbol{\Sigma}_{M} = \left[\frac{\partial \mathcal{H}}{\partial \boldsymbol{\theta}} \boldsymbol{\Sigma} \frac{\partial \mathcal{H}}{\partial \boldsymbol{\theta}}^{T}\right]$$

Proof: Use the conclusion in Theorem 4.0.11 and then apply the delta method to the differentiable map $\mathcal{H}$. ■

Remark It should be pointed out that the MOM estimation is better suited to estimate the parameters in VIX_t^2 as there are fewer (and simpler) moments to compute. But Kagan and Nagaev [43] showed that moment-based estimators require substantial amounts of data if one wants to consistently estimate more than two population moments simultaneously. If we take into account the noise contained in the high frequency data , moments based methods are probably not a good choice for (jointly) estimating the parameters.

Remark As mentioned in the beginning of this chapter, if one is only interested in the parameters $(\boldsymbol{\eta}, \rho)$ in the BN-S model[1], *Realized Quadratic Varia-*

[1]Here λ can still be estimated by the autocorrelation functions of the VIX^2 data

tion and the more general *Multipower Variation* can be used to aid the estimation. Recall we observe R_t over equi-spaced partition $\pi_n = \{T_{n,0}, \ldots, T_{n,n}\}$ with $\max_{1 \le k \le n}\{T_{n,i} - T_{n,i-1}\} \to 0$ as $n \to \infty$. Let $\Delta_n = T/n$. The normed p-th power variation proposed by Barndorff-Nielsen and Shephard (cf [9], [10], [11] and the references in there) is defined as

$$\Delta_n^{\gamma} V_p(X, \pi_n) = \sum_{i=1}^{n} \Delta_n^{\gamma} \left| R_{T_{n,i}} - R_{T_{n,i-1}} \right|^p.$$

A further extension is the normed p-th bipower variation:

$$V_{r,s}(X, \pi_n, \Delta_n^{\gamma}, \Delta_n^{\delta}) = \sum_{i=1}^{n} \Delta_{n,i+1}^{\gamma} \left| R_{T_{n,i+1}} - R_{T_{n,i}} \right|^r \Delta_{n,i}^{\delta} \left| R_{T_{n,i}} - R_{T_{n,i-1}} \right|^s.$$

The purpose of constructing various realized (bi)power variations is to study the quadratic variation of the return process R_t, whose form under the BN-S model is given by

$$[R, R]_t = \int_0^t \sigma_s^2 ds + \rho^2 \sum_{s \le t} (Z_s - Z_{s-})^2.$$

Woerner [81] showed

$$\Delta_n^{1-\frac{p}{2}} V_2(R, \pi_n)\, \mu_2^{-1} \xrightarrow{\mathcal{P}} \int_0^h \sigma_s^s \,\mathrm{d}s \;+ \sum \rho^2 \big((R_s - R_{s-})^2; 0 < s \le h\big)$$

$$\Delta_n^{1-\frac{p}{2}} V_{r,s}(R, \pi_n)\, \mu_r^{-1}\, \mu_s^{-1} \xrightarrow{\mathcal{P}} \int_0^h \sigma_s^s \,\mathrm{d}s$$

for μ_r denoting the $r-th$ absolute moment of a standard normal r.v.. By increasing the data sampling frequency, i.e., letting $\Delta_n \to 0$, $\int_0^t \sigma_s^2 ds$ and the sum squared-jumps $\rho^2 \sum_{s\leq t}(Z_s - Z_{s^-})^2$ can be approximated with high accuracy. Therefore one can use those realized variations as their limiting counterparts, that is, we can assume $\int_0^t \sigma_s^2 ds$ and $\rho^2 \sum_{s\leq t}(Z_s - Z_{s^-})^2$ are actually observed, and one can choose proper schemes to find the estimators for $(\boldsymbol{\eta}, \rho)$. For further discussion, check [8], [34], [42], [73], [78] and [80] for details.

5. DISCUSSION AND FUTURE STUDY

In this thesis, we explore the applicability of the well studied Minimum Disparity method for performing parameter estimation to the BN-S stochastic volatility model. By proving the bivariate series (X_i, σ_i^2) implied by the BN-S model to be geometric ergodic with smooth stationary distribution, we analyze the limiting properties of various estimators. In particular, we demonstrate how to combine the S&P 500 data and the VIX data to consistently estimate the parameters in the Γ-OU BN-S model using the Negative Exponential disparity estimator. Consistency and asymptotic normality of the NEDE are proved under relatively weak conditions. By using the geometric ergodicity again and verifying the finiteness of the moments, strong consistency and asymptotic normality of the MOM estimator are proved under the Γ-OU and the IG-OU BN-S model, provided that both X_i and σ_i^2 are observed. Although this conclusion is not directly applicable to empirical studies, but one can still use the geometric ergodicity of the X_i to study other estimation schemes based on functions of X_i.

In the process of this investigation, we found new problems arose from different aspects of the study, for example, conceptual understandings, technical difficulties, methodology issues and implementation challenges. Here we list a couple of topics which we think deserved a closer examination in the future study.

(1) Numerical implementation of the NEDE. Although there is no data analysis included in this study, from some trial simulations I find that it is possible for the density $m_{\boldsymbol{\theta}}(x)$ implied by the Γ-OU BN-S model to have similar shape for different sets of parameters. This suggests highly accurate and stable numeric methods are required in order to produce consistent estimates for all parameters simultaneously. The simulated annealing method used in Takada [70] does seem to be a good candidate, however, one should keep the dimension issue in mind.

(2) Robustness of NEDE under dependent data. We have yet to produce discussions over the trade-off between robustness and efficieny when applying the NEDE to the Γ-OU BN-S model. This is partly due to the lack of a proper notion for influence function under the jump diffusion model setting. As the traditional influence function theory considers how single contamination affects an observation from an i.i.d. set of data, we need to consider the effect of an outlier over all observations jointly. After some literature reviews, the pioneering work by Martin and Yohai [51] who gave a general framework for influence function over time series and the recent study by Toronjadze [75] who investigated influence function on stochastic equations for semimartingale seem to be the right approach to define a concrete definition for influence function to the BN-S model.

(3) Check the model goodness-of-fit. In this paper the disparity (deviation) concept is used to drive the parameter estimation, but its classical role is to analyze the

goodness-of-fit of the given model. A proper goodness-of-fit test statistics for our model should be derived to accompany the discussion of robustness and efficicency.

(4) Perform Taylor expansion for other BN-S models. We use a very ad-hoc method to justify the Taylor expansion for the Γ-OU BN-S model in Appendix B.2. But it is no doubt that those steps are hard to be reproduced when the joint distribution of increment processes (see (B.11)) is unknown. However, as the moment bounds results (Proposition B.2.9 and Proposition B.2.10) in principle hold for other BN-S models thanks to the Gaussian component, we can justify the Taylor expansion for other models by showing the tails of those derivatives grow at most in a polynomial order of $|x|$. Since the characteristic function of X_i and its derivatives can be derived explicitly, a method to link the tail behavior of functions to their Fourier (Laplace) transforms will help to solve this problem.

(5) Extend the functional delta method. We find the functional delta method to be a very convenient tool to study estimators which are functionals of the kernel density estimate. Although one needs advanced functional analysis skills to study various functional derivatives, compared the steps between Lemma 3.2.8 and those in Appendix B.3, one finds the central limit theorem can be directly applied without passing the proof from Pearson residuals to Hellinger residuals.

(6) Model selection by using disparity. If we are able to extend the NEDE to different families of stochastic volatility models, then we can use the disparity

as a quantitative measure to choose the model with the best fit. It will be interesting to compare such measure to the classical AIC and BIC under different circumstances.

(7) Consider disparities between other densities. As nonparametric estimates for characteristic functions, spectral densities and Lévy densities have been well studied, we can estimate parameters by minimizing appropriate distances between these functions.

APPENDIX

A. LEMMAS AND FACTS IN CHAPTER 2

A.1 Important Lemmas

Let X be a Markov chain defined on the sample space $\mathcal{S}$ and φ is a σ-finite measure defined on $\mathcal{B}(\mathcal{S})$. Following are several useful results related to the study of stability of Markov chain.

Lemma A.1.1: **(Tuominen and Tweedie 1979, Proposition 1.2)**. If the transition operator $(\mathcal{P}_t)_{t\in\mathbb{R}_+}$ for a Markov process X is simultaneously φ-irreducible, then any Δ-skeleton chain of X is aperiodic.

Lemma A.1.2: **(Meyn and Tweedie 1992, Theorem 3.4 (ii); Meyn and Tweedie 2009, Theorem 5.5.7)** Suppose X is φ-irreducible and aperiodic. If X has the Feller property and suppφ has non-empty interior, then all compact sets of $\mathcal{S}$ are small.

A.2 Exponential Ergodicity of univariate OU Process

Note: the result quoted below is the one-dimensional version of the original theorem in [52].

Lemma A.2.1: (**Masuda 2004, Theorem 4.3**) Let λ be positive and X be the strictly stationary OU process given by

$$X_t = e^{-\lambda t} X_0 + \int_0^t e^{-\lambda(t-s)} dZ_s, \qquad t \in \mathbb{R}^+$$

with a self-decomposable marginal distribution F. If we have

$$\int |x|^p F(\mathrm{d}x) < \infty \tag{A.1}$$

for some $p > 0$, then there exists a constant $a > 0$ such that $\beta_X(t) = O(e^{-at})$ as $t \to \infty$. In particular, X is ergodic.

Since the proof of our Theorem 2.4.3 is essentially the same as Masuda's proof to the above lemma. we excerpt the original proof from Masuda for reader's reference. Some notations are slightly modified to be consistent with our discussion.

Proof: Let $\mathbb{N} = \{1, 2, 3, \ldots\}$, then for each Δ one has

$$X_n^\Delta = e^{-\lambda\Delta} X_{n-1}^\Delta + \xi_n,$$

where $\xi = (\xi_n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed random variables with marginal law $\mathcal{L}(\xi_1) = \mathcal{L}(\int_0^\Delta e^{-\lambda(\Delta-s)} dZ_s)$. It is easy to see that X^Δ

is also strictly stationary with the same marginal distribution F as X.

Ergodic with Geometric Mixing rate. First, the author shows that X^{Δ} is geometrically ergodic. Let S_F denote the support of F, one has $\lim_{n\to\infty} P_{n\Delta}(A, x) = \lim_{t\to\infty} P_t(x, A) = F(A)$ for any $\Delta > 0$ and $A \in \mathcal{B}(S_F)$. Thus $X^{(\Delta)}$ is simultaneously F-irreducible. Hence by Lemma A.1.1, $X^{(\Delta)}$ is aperiodic for any Δ.

Without loss of generality, assume $p \in (0, 1]$. Put $\delta = |e^{-\lambda\Delta}|$. Under condition (A.1), $E[|X_1|^p] < \infty$, thus we will verify (2.8) and (2.9) for function $g(y) = |y|^p$. Since we restrict λ to be strictly positive, then $\delta < 1$ for positive Δ. Fix this choice of Δ for the rest of the proof.

From the strict stationarity of X^{Δ}, one has

$$
\begin{aligned}
E[|\xi_1|^p] &= E[|X_1^{\Delta} - e^{-\lambda\Delta} X_0^{\Delta}|^p] \\
&\leq E[(|X_1^{\Delta}| + \delta|X_0^{\Delta}|)^p] \\
&= (1 + \delta^p)E[|X_0^{\Delta}|^p] < \infty
\end{aligned}
$$

Put $C_\eta = \{x \in S_F : |x| \leq \eta\}$ for some constant $\eta > 0$; then C_η is a small set since it is compact. Denote its complement as C_η^c. Then, since the support of F is unbounded, so for any η the set C_η^c is not empty. As $X_0^{\Delta} = X_0$ is chosen to be independent of Z_t, so X_0^Δ is also independent of ξ_1, and one has

$$\int_{C_\eta^c} |y|^p P_\Delta(dy, x_0) \le E[|e^{-\lambda\Delta}x_0 + \xi_1|^p]$$

$$\le \delta^p \eta^p + E[|\xi_1|^p] < \infty$$

for any $x_0 \in C_\eta$. Since this upper bound does not depend on x_0, then (2.8) is obtained. On the other hand, for $x_1 \in C_\eta^c$, let c_1 be a constant such that $\delta < c_1 < 1$. Then,

$$\begin{aligned}\int_{C_\eta} |y|^p P_\Delta(dy, x_1) &\le E[|e^{-\lambda\Delta}x_1 + \xi_1|^p] \\ &\le \delta|x_1|^p + E[|\xi_1|^p] \\ &= c_1|x_1|^p - \big((c_1 - \delta)|x_1|^p - E[|\xi_1|^p]\big) \\ &= c_1|x_1|^p - c_2\end{aligned}$$

Since $E[|\xi_1|^p]$ is finite, one can choose η large enough so that $c_2 > 0$. So we obtain the bound (2.9), hence from Proposition 2.4.1 we concludes that X^Δ is geometrically ergodic.

Exponential Mixing rate. From the conclusion of step 1, there exists a constant ρ such that $\rho \in (0, 1)$ and

$$\int \sup_{|f|\le 1} |\mathcal{P}_{n\Delta} f(x) - F(f)| F(\mathrm{d}x) = O(\rho^n), \quad \text{as } n \to \infty \tag{A.2}$$

Denote by $[t]$ the integer part of $t \in \mathbb{R}_+$, and let $t_\Delta = [t/\Delta]\Delta$ and $f_t = \mathcal{P}_{t_\Delta} f \in b\mathcal{B}(S_F)$. Then using the property of semigroup, the invariance of F and (A.2) yield that

$$
\begin{aligned}
\beta_X t &= \int \sup_{|f|\leq 1} |\mathcal{P}_t f(x) - F(f)| F(\mathrm{d}x) \\
&= \int \sup_{|f|\leq 1} |[\mathcal{P}_{t_\Delta} \mathcal{P}_{t-t_\Delta} f](x) - F(f)| F(\mathrm{d}x) \\
&= \int \sup_{|f|\leq 1} |[\mathcal{P}_{t_\Delta} \mathcal{P}_{t-t_\Delta} f](x) - F(\mathcal{P}_{t-t_\Delta} f)| F(\mathrm{d}x) \\
&\leq \int_{\mathbb{R}^2} \sup_{|f_t|\leq 1} |\mathcal{P}_{t_\Delta} f_t(x) - F(f_t)| F(\mathrm{d}x) \\
&= O(\rho^{t_\Delta/\Delta})
\end{aligned}
$$

as $t \to \infty$, so by taking $a = -(\log \rho)/\Delta$ we complete the proof. ■

B. RESULTS, DERIVATIONS AND EXTENSIONS OF MDE

B.1 Efficiency, Robustness and Asymptotic Properties of MDE

Here we present the results by Basu and Lindsay which are related to the efficiency and robustness of the MDE. The implications of these results have been previously discussed in Section 3.1. We begin by defining some expressions (where notations have been adapted to be consistent with the current discussion).

Let ∂_j and ∂_{jk} represent the partial derivatives with respect to θ_j and θ_j, θ_k and write $\tilde{u}_j(x,\theta) = \partial_j \ln m^*_\theta(x)$ and $\tilde{u}_{jk}(x,\theta) = \partial_{jk} \ln m^*_\theta(x)$. Assuming that one can interchange the order of differentiation and integration, let

$$u^*_j(x,\theta) = \int k(x;t,h)\tilde{u}_j(x,\theta)\,\mathrm{d}x \;=\; \partial_j \int \ln m^*_\theta(x) k(x;t,h)\,\mathrm{d}x\,,$$

$$u^*_{jk}(x,\theta) = \int k(x;t,h)\tilde{u}_{jk}(x,\theta)\,\mathrm{d}x \;=\; \partial_{jk} \int \ln m^*_\theta(x) k(x;t,h)\,\mathrm{d}x\,.$$

Let the $p \times p$ matrix $J^*(\theta)$ be defined as the information matrix corresponding to a random variable with pdf $m^*_\theta(x)$, with its jk-th element is given by $E_\theta[-u^*_{jk}(X,\theta)]$. Let $s^*(x) = \int k(x;t,h)s(t)\,\mathrm{d}t$ be the kernel smoothed version of $s(x)$. Recall θ^s is defined as the minimizer of

$$\rho(s^*, m_\theta^*) \;=\; \int G\Big(\tfrac{s^*(x)}{m_\theta^*(x)} - 1\Big)\, m_\theta^*(x)\,\mathrm{d}x\,.$$

Let $\delta_s^*(x) = s^*(x)/m_{\theta^s}^*(x) - 1$. Define $J^{*s}(\theta^s)$ to be the $p \times p$ matrix whose jk-th element is given by

$$\int A'(\delta_s^*)\tilde{u}_j(x,\theta^s)\tilde{u}_k(x,\theta^s)s^*(x)\,\mathrm{d}x \;-\; \int A'(\delta_s^*)\nabla_{jk} m_{\theta^s}^*(x)\,\mathrm{d}x$$

and let $v^*(t,\theta^s)$ be the p-dimensional vector whose j-th component is

$$\int A'(\delta_s^*)\tilde{u}_j(x,\theta^s)k(x;t,h)\,\mathrm{d}x \;-\; \int A'(\delta_s^*)\tilde{u}_j(x,\theta^s)s^*(x)\,\mathrm{d}x\,.$$

Lemma B.1.1: **(Basu and Lindsay 1994, Lemma 5.1)** Let $S(x)$ be the true distribution which is not necessarily in the family of model $\{m_\theta^*(x)\}$. For the minimum disparity functional T, let $T(S) = \theta^s$. Then the influence function of T (see (3.2)) has the form $T'(y) = [J^{*s}(\theta^s)]^{-1}v^*(y,\theta^s)$. If $S = M_{\theta_0}$ for some θ_0, then the above reduces to $T'(y) = [J^*(\theta_0)]^{-1}u^*(y,\theta_0)$. If in addition k is a transparent kernel for the family M_θ then we get $T'(y) = [I(\theta_0)]^{-1}u(y,\theta_0)$, where $I(\theta)$ is the Fisher information about θ in m_θ.

Lemma B.1.2: **(Basu and Lindsay 1994, Theorem 5.1)** Let $T''(y) = \frac{\partial^2}{\partial\epsilon^2}T(M_{\theta,\epsilon})|_{\epsilon=0}$. Then for an estimating function of the type $\int A(\delta^n)\nabla m_\theta^*(x)\,\mathrm{d}x$, we have

$$T''(y) = T'(y)\Big[\int \tilde{u}^2(x,\theta)m_\theta^*(x)\,\mathrm{d}x\,\Big]^{-1}\big\{f_1(y) + A_2 f_2(y)\big\},$$

where

$$f_1(y) = 2\nabla u^*(x,\theta) - 2E_\theta\big[\nabla u^*(X,\theta)\big] + T'(y)E_\theta\big[\nabla^2 u^*(X,\theta)\big],$$

and

$$f_2(y) = [u^*(X,\theta)]^{-1}\Big[\int \tilde{u}^2(x,\theta)m_\theta^*(x)\,\mathrm{d}x\Big]\Big[\int k^2(x;y,h)\tilde{u}(x,\theta)(m_\theta^*(x))^{-1}\,\mathrm{d}x\Big]$$
$$-2\int k(x;y,h)\tilde{u}^2(x,\theta)\,\mathrm{d}x + T'(y)\int \tilde{u}^3(x,\theta)m_\theta^*(x)\,\mathrm{d}x$$

Second, we present the result which considers the consistency and asymptotic normality of the MDE. Again, we need to introduce some definitions.

Definition

- The kernel integrated family of distributions is *smooth* if the conditions of Lehmann and Casella ([47], pp.440-441) are satisfied with $m_\theta^*(x)$ in place of $f(x|\theta)$. Under those conditions, $m_\theta^*(x)$ is required to have a certain degree of integrability and differentiability with respect to both x and θ. Also, the Fisher information matrix based on $m_\theta^*(x)$ needs to be finite.

- The true density $s(x)$ is *compatible* with $m_\theta(x)$ if $s(x) > 0$ on the common support of $m_\theta(x)$ and the functions $M_{jkl}(x)$, $M_{jk,l}(x)$, $M_{j,k,l}(x)$ have finite expectations with respect to $s^*(x)$; in addition (B.1) holds and the integrals $\int(s^*)(x)^{1/2}|\tilde{u}_j(x)\tilde{u}_k(x)|\,\mathrm{d}x$ and $\int(s^*)(x)^{1/2}|\tilde{u}_{jk}(x)|\,\mathrm{d}x$ are finite for all j and k.

Lemma B.1.3: **(Basu and Lindsay 1994, Theorem 6.1)** Suppose that the conditions

$$|\tilde{u}_{jkl}(x)| \leq M_{jkl}(x), \quad |\tilde{u}_{jk}(x)\tilde{u}_l(x)| \leq M_{jk,l}(x), \quad |\tilde{u}_j(x)\tilde{u}_k(x)\tilde{u}_l(x)| \leq M_{j,k,l}(x)$$

hold for all j, k and l, for all θ in a neighborhood B_a of θ^s, where $M_{jkl}(x)$, $M_{jk,l}(x)$ and $M_{j,k,l}(x)$ have finite expectations with respect to $m^*_\theta(x)$ for all $\theta \in B_a$. Assume that the residual adjustment function $A(\delta)$ corresponding to a particular disparity measure ρ is regular, m_θ is smooth, $s(x)$ is compatible with m_θ and the matrix $J^{*s}(\theta^*)$, as defined in Lemma B.1.1 is positive definite. Then there exists a consistent sequence of roots θ_n to the minimum disparity estimating equations. The asymptotic distribution of $n^{1/2}(\theta_n - \theta^s)$ is MVN with mean 0 and variance $[J^{*s}(\theta^s)]^{-1}V_s[J^{*s}(\theta^s)]^{-1}$ where V_s is the quantity V in (B.1) evaluated at $\theta = \theta^s$.

Remark Basu and Lindsay did not provide a detailed proof of this theorem in their paper and they pointed to [49] and [65] for further reference. After carefully examining the proof in the referred literature, we believe an assumption on the integrability of $A(\delta)$ should also be included in the assumptions for completeness. However, the authors actually assumed such integrability conditions implicitly when deriving the minimum disparity estimating equations (see equation (2.6) in [12]).

We shall follow the arguments by Lehmann and Casella ([47] Chapter 6, Theorem 5.1) and Lindsay ([49] Theorem 33) to produce a heuristic proof of Lemma B.1.3. This helps to identify the sufficient conditions and their roles in proving the consistency and asymptotic normality of the general MDE. Besides, we would like to find out the extra conditions needed when the data are dependent and the Negative Exponential disparity is used.

In what follows, $f^*(x)$ is the kernel density estimate computed based on n i.i.d. data $\{x_i\}$. First, let us present several lemmas from [12] and discuss their consequences.

(B-L1). ([12] Lemma 6.1, Lemma 6.2) $n^{1/4}(f^{*1/2}(x) - s^{*1/2}(x)) \to 0$ with probability 1 if $\lambda(x) < \infty$ where

$$\lambda(x) = \int k^2(x; t, h) s(t) dt - [s^*(x)]^2.$$

(B-L2). ([12] Lemma 6.3 (i), Lemma 6.4, Lemma 6.5) If

$$\int s^{*1/2}(x) |\nabla \ln m_\theta^*(x)| \, \mathrm{d}x \; < \infty, \tag{B.1}$$

then

$$\int n^{1/2} \big[A(\delta^n(x)) - A(\delta_s^*(x)) \big] \nabla m_\theta^*(x) \, \mathrm{d}x$$

and

$$\int n^{1/2} (\delta^n(x) - \delta_s^*(x)) A'(\delta_s^*(x)) \nabla m_\theta^*(x) \, \mathrm{d}x$$

are asymptotically equivalent. That is,

$$E \left| \int n^{1/2} \Big[A(\delta^n(x)) - A(\delta_s^*(x)) - (\delta^n(x) - \delta_s^*(x)) A'(\delta_s^*(x)) \Big] \nabla m_\theta^*(x) \, \mathrm{d}x \right| \to 0$$

(B-L3). ([12] Lemma 6.3 (ii), Corollary 6.1) Suppose that

$$V = \mathrm{Var}\Big(\int k(x, X, h)\, A'(\delta_s^*(x))\, \nabla \ln m_\theta^*(x)\, \mathrm{d}x\Big) \tag{B.2}$$

is finite, using the result in (1), Basu and Lindsay showed

$$n^{1/2} \int \big(\delta^n(x) - \delta_s^*(x)\big)\, A'(\delta_s^*(x))\, \nabla m_\theta^*(x)\, \mathrm{d}x \;\to\; N(0, V).$$

for a regular RAF $A(\cdot)$. From the asymptotic equivalence shown in $(B-L2)$, one has

$$n^{1/2} \int \Big[A(\delta^n(x)) - A(\delta_s^*(x))\Big] \nabla m_\theta^*(x)\, \mathrm{d}x \;\to\; N(0, V). \tag{B.3}$$

Result (B.3) implies the un-normalized integral converges to 0 as $n \to \infty$. This fact will be used in the study of consistency and of MDE.

A heuristic proof of Lemma B.1.3: Recall θ^s is the unique minimizer of the disparity $\rho(s^*(x), m_\theta^*(x))$, that is, it solves

$$\nabla \rho(s^*(x), m_\theta^*(x)) = \nabla \int G(\delta_s^*(x)) m_\theta^*(x)\, \mathrm{d}x \; = 0.$$

where $\delta_s^*(x) = (s^*(x) - m_{\theta^s}^*(x))/m_{\theta^s}^*(x)$. Also, let θ_n be the solution to the minimum disparity equation

$$\nabla \int G(\delta^n(x)) m_\theta^*(x)\, \mathrm{d}x \; = 0$$

for each n where $\delta^n(x) = (f^*(x) - m^*_\theta(x))/m^*_\theta(x)$. As suggested in [47], to prove the local consistency of θ_n, one considers the Taylor expansion of $\rho(f^*(x), m^*_\theta(x))$ on θ in a small p-dimensional neighborhood B_a of θ^s with radius a. Recall $\partial_i = \frac{\partial}{\partial\theta_i}$ and $\partial_i^{n_i} = \frac{\partial}{\partial\theta_i^{n_i}}$. Then,

$$\begin{aligned}\rho(f^*, m^*_\theta) = \rho(f^*, m^*_{\theta_s}) &+ \sum_{i=1}^{p} \partial_i \rho(f^*, m^*_\theta)|_{\theta=\theta^s}(\theta_i - \theta_i^s) \\ &+ \frac{1}{2}\sum_i\sum_j \partial_{ij}\rho(f^*, m^*_\theta)|_{\theta=\theta^s}(\theta_i - \theta_i^s)(\theta_j - \theta_j^s) \\ &+ \frac{1}{6}\sum_{n_1+\cdots+n_p=3} \frac{\partial^3}{\partial_1^{n_1}\ldots\partial_p^{n_p}}\rho(f^*, m^*_\theta)|_{\theta=\theta^s}\cdot\frac{(\theta_1-\theta_1^s)^{n_1}\ldots(\theta_p-\theta_p^s)^{n_p}}{n_1!\ldots n_p!} \\ &+ o_p(a^4)\end{aligned}$$

With a slight abuse of notation, we shall use $m^*_{\theta^s}(x)$ and $\partial_i m^*_{\theta^s}(x)$ to represent $m^*_\theta(x)$ and $\partial_i m^*_\theta(x)$ evaluated at $\theta = \theta^s$ respectively. Notations for the higher order derivatives will be understood similarly. We want to study asymptotic behaviors of the terms in the above expansion. To apply the steps in Lehmann and Casella, one needs to show, as $n \to \infty$, the first derivatives of $\rho(f^*, m^*_\theta)$ with respect to θ converge to 0, the matrix of second derivatives converges to a non-negative definite matrix, and all the third derivatives converge to some finite quantities.

(1). Coefficients of the Linear Terms: $\partial_i\rho(f^*, m^*_\theta)$

If differentiation and integration can be interchanged, one has

$$
\begin{aligned}
&\partial_i \rho(f^*, m_\theta^*)|_{\theta=\theta^s} \\
&= \int_{\mathbb{R}} \partial_i \Big[G(\delta^n(x)) m_{\theta^s}^*(x) \Big] \, \mathrm{d}x \\
&= \int_{\mathbb{R}} \Big[G'(\delta^n(x)) \frac{f^*(x)}{-m_{\theta^s}^*(x)^2} \cdot \partial_i m_{\theta^s}^*(x) \cdot m_{\theta^s}^*(x) + G(\delta^n(x)) \, \partial_i m_{\theta^s}^*(x) \Big] \, \mathrm{d}x \\
&= -\int_{\mathbb{R}} A(\delta^n(x)) \, \partial_i m_{\theta^s}^*(x) \, \mathrm{d}x \, .
\end{aligned}
$$

It is clear that the condition needed for differentiating under the integral sign is:

$$
\int_{\mathbb{R}} \sup_{\theta \in \boldsymbol{B}_a} \left| A(\delta^n(x)) \, \partial_i m_\theta^*(x) \right| \mathrm{d}x \ < \infty \tag{B.4}
$$

To show the linear term converges to 0 as $n \to \infty$ for $\theta \in \boldsymbol{B}_a$, it suffices to show

$$
\int_{\mathbb{R}} \big[A(\delta^n(x)) - A(\delta_s^*(x)) \big] \, \partial_i m_{\theta^s}^*(x) \, \mathrm{d}x \xrightarrow{\mathcal{P}} 0 \qquad \text{as } n \to \infty,
$$

because

$$
\textstyle\int_{\mathbb{R}} A(\delta_s^*(x)) \, \partial_i m_{\theta^s}^*(x) \, \mathrm{d}x \ = \ \partial_i \rho(s^*, m_{\theta^s}^*) \ = \ 0
$$

for all i by the definition of θ^s. From the result of (B.3), we know the above convergence is true for all i. Therefore, the coefficients of the linear terms in the Taylor expansion converge to 0.

(2). Coefficients of the Quadratic Terms: $\partial_{ij} \rho(f^*, m_\theta^*)$

Let $\tilde{u}_i(x, \theta) \triangleq \partial_i \ln m_\theta^*(x)$. If differentiation and integration can be interchanged, one has:

$$\begin{aligned}
\partial_{ij}\rho(f^*, m_\theta^*)|_{\theta=\theta^s} &= \int_{\mathbb{R}} \partial_{ij}\Big[G(\delta^n(x))m_{\theta^s}^*(x)\Big]\,\mathrm{d}x \\
&= \int_{\mathbb{R}} A'(\delta^n(x))\frac{f^*(x)}{m_{\theta^s}^*(x)^2}\cdot\,\partial_i m_{\theta^s}^*(x)\cdot\,\partial_j m_{\theta^s}^*(x)\,\mathrm{d}x - \int_{\mathbb{R}} A(\delta^n(x))\,\partial_{ij}m_{\theta^s}^*(x)\,\mathrm{d}x \\
&= \int_{\mathbb{R}} A'(\delta^n)(\delta^n(x)+1)\cdot\tilde{u}_i(x,\theta^s)\cdot\tilde{u}_j(x,\theta^s)\cdot m_{\theta^s}^*(x)\,\mathrm{d}x - \int_{\mathbb{R}} A(\delta^n)\,\partial_{ij}m_{\theta^s}^*(x)\,\mathrm{d}x
\end{aligned}$$

This suggests the following conditions are needed:

$$\begin{aligned}
&\int_{\mathbb{R}} \sup_{\theta\in B_a}\Big|A'(\delta^n)(\delta^n(x)+1)\cdot\tilde{u}_i(x,\theta)\cdot\tilde{u}_j(x,\theta)\Big|m_{\theta^s}^*(x)\,\mathrm{d}x \\
=&\int_{\mathbb{R}} \sup_{\theta\in B_a}\Big|A'(\delta^n)\cdot\tilde{u}_i(x,\theta)\cdot\tilde{u}_j(x,\theta)\Big|f^*(x)\,\mathrm{d}x \;<\infty
\end{aligned}$$

and

$$\int_{\mathbb{R}} \sup_{\theta\in B_a}\Big|A(\delta^n)\,\partial_{ij}m_{\theta^s}^*(x)\Big|\,\mathrm{d}x \;<\infty \tag{B.5}$$

Since $f^*(x) \xrightarrow{\mathcal{P}} s^*(x)$, we can instead assume the last inequality holds with $f^*(x)$ replaced by $s^*(x)$. Notice that $A'(\delta)$ is bounded because $A(\delta^n)$ assumed to be regular, the first condition simplifies to

$$\int_{\mathbb{R}} \sup_{\theta\in B_a}\Big|\tilde{u}_i(x,\theta)\cdot\tilde{u}_j(x,\theta)\Big|s^*(x)\,\mathrm{d}x \;<\infty. \tag{B.6}$$

Under (B.5) and (B.6), $\partial_{ij}\rho(f^*, m_\theta^*)|_{\theta=\theta^s}$ converges to

$$\int_{\mathbb{R}} A'(\delta_s^*(x))\tilde{u}_i(x,\theta^s)\cdot\tilde{u}_j(x,\theta^s)s^*(x)\,\mathrm{d}x - \int_{\mathbb{R}} A(\delta_s^*)\,\partial_{ij}m_{\theta^s}^*(x)\,\mathrm{d}x \tag{B.7}$$

in probability as $n \to \infty$. We also need to assume that the matrix J where its ij-th element given by the right-hand side of (B.7) is a positive definite matrix.

(3). Coefficients of the Cubic Terms: $\partial_{ijk}\rho(f^*, m_\theta^*)$ or $\partial_{ij,k}\rho(f^*, m_\theta^*)$

There are four types of cubic terms in the expansion and the computations become quite involved. We skip the details and present the expressions after the integrand has been differentiated. We also substitute $f^*(x)$ by $s^*(x)$ in the final form of the conditions.

- For terms like $\int_{\mathbb{R}} A''(\delta^n)\frac{f^*(x)^2}{m^*_{\theta^s}(x)^4} \cdot \partial_i m^*_{\theta^s}(x) \cdot \partial_j m^*_{\theta^s}(x) \cdot \partial_k m^*_{\theta^s}(x)\,\mathrm{d}x$: recall $\delta^n = f^*(x)/m^*_\theta(x) - 1$, rewrite this integral as

$$\int_{\mathbb{R}} \left[A''(\delta^n) f^*(x)(\delta^n + 1)\right]\frac{1}{m^*_{\theta^s}(x)^3} \cdot \partial_i m^*_{\theta^s}(x) \cdot \partial_j m^*_{\theta^s}(x) \cdot \partial_k m^*_{\theta^s}(x)\,\mathrm{d}x$$

  Since $A''(\delta^n)(\delta^n + 1)$ is bounded because $A(\delta)$ is regular, one needs to assume

$$\int_{\mathbb{R}} \sup_{\theta \in \tilde{B}_a} \left|\tilde{u}_i(x,\theta) \cdot \tilde{u}_j(x,\theta) \cdot \tilde{u}_k(x,\theta)\right| s^*(x)\,\mathrm{d}x \ < \infty \tag{B.8}$$

- For terms like $\int_{\mathbb{R}} A'(\delta^n)\frac{f^*(x)}{m^*_{\theta^s}(x)^3} \cdot \partial_i m^*_{\theta^s}(x) \cdot \partial_j m^*_{\theta^s}(x) \cdot \partial_k m^*_{\theta^s}(x)\,\mathrm{d}x$: since $A'(\delta^n)$ is bounded, one needs (B.8).

- For terms like $\int_{\mathbb{R}} A'(\delta^n)\frac{f^*(x)}{m^*_{\theta^s}(x)^2} \cdot \partial_{ij} m^*_{\theta^s}(x) \cdot \partial_k m^*_{\theta^s}(x)\,\mathrm{d}x$: one needs to assume

$$\int_{\mathbb{R}} \sup_{\theta \in \tilde{B}_a} \left|\frac{\partial_{ij} m^*_{\theta^s}(x)}{m^*_{\theta^s}(x)} \cdot \frac{\partial_k m^*_{\theta^s}(x)}{m^*_{\theta^s}(x)}\right| s^*(x)\,\mathrm{d}x \ < \infty \tag{B.9}$$

- At last,

$$\int_{\mathbb{R}} \sup_{\theta \in \tilde{B}_a} \left| A(\delta^n)\, \partial_{ijk} m^*_{\theta^s}(x) \right| \mathrm{d}x \ < \infty \tag{B.10}$$

If condition (B.4) to (B.10) (except (B.7)) hold for all i, j and k less than p, then as n gets large we can show the coefficients of the linear terms are of order $O_p(a^2)$ while the coefficients of the quadratic and cubic terms are of order $O_p(1)$. This implies the leading order terms in the expansion are the quadratic ones with order $O_p(a^2)$. Then,

$$\min_{\theta \in B_a} \left(\rho(f^*, m^*_\theta) - \rho(f^*, m^*_{\theta^s}) \right) > 0$$

with probability converging to 1. Therefore we know the disparity $\rho(f^*, m^*_\theta)$ has a local minimum in B_a and the minimizer $\theta_n \in B_a$ for any $a > 0$ when n is sufficiently large. This proves the consistency of θ_n.

Once the consistency of θ_n is obtained, one can prove its asymptotic normality by performing the Taylor expansion on $\sqrt{n} \nabla \int \rho(f^*(x), m^*_{\theta_s}(x))\, \mathrm{d}x$ with respect to θ and use the result in (B.3). ■

Remark The above derivation is the generalization of the proof used in Section 3.2. The main difference is assumption (B.4), (B.5) and (B.10) which involve the boundedness (or integrability) of $A(\delta)$.

B.2 Deriving the Taylor Expansion of $\rho(f^*(x), m^*_{\boldsymbol{\theta}}(x))$ with respect to $\boldsymbol{\theta}$ in the Γ-OU BN-S Model

In this section, a detailed derivation of the Taylor expansion (3.9) discussed in Section 3.2 is provided. Simliar notations previously defined in Section 3.2 will be used here unless stated otherwise.

To begin with, the density of $m_{\boldsymbol{\theta}}(x)$ and its derivatives will be derived as they will be repeatedly used in this section. Recall $m_{\boldsymbol{\theta}}(x)$ is the stationary density of X_i defined in the BN-S model (equation (2.1))[1]:

$$X \,\Big|\, \textstyle\int_0^h \sigma_t^2 \,\mathrm{d}t \,, \int_0^h \,\mathrm{d}Z_t \;\sim\; N\Big(\mu + \beta \int_0^h \sigma_s^2 \,\mathrm{d}t \,+ \rho \int_0^h \,\mathrm{d}Z_t, \int_0^h \sigma_t^2 \,\mathrm{d}t \,\Big)$$

Notice that

$$\textstyle\int_0^h \sigma_t^2 \,\mathrm{d}t \, = (1 - e^{-h})\sigma_0^2 + \int_0^h (1 - e^{-h+t}) \,\mathrm{d}Z_t$$

So if we denote (suppressing the notation h in the names of the r.v.'s)

$$S = \sigma_0^2, \qquad Y = \int_0^h (1 - e^{-h+t}) \,\mathrm{d}Z_t, \qquad \text{and} \qquad W = \int_0^h \,\mathrm{d}Z_t, \tag{B.11}$$

then the density of X is given by the following expectation:

$$m_{\boldsymbol{\theta}}(x) = E\Big[\frac{1}{\sqrt{2\pi((1 - e^{-h})S + Y)}} \exp\Big(- \frac{(x - \mu - \beta((1 - e^{-h})S + Y) - \rho W)^2}{2((1 - e^{-h})S + Y)}\Big)\Big] \tag{B.12}$$

[1] the subscript i will be suppressed

Since S is independent of (Y, W) and follows a Gamma(ν, α) distribution, so to find the joint density of (S, Y, W), we just need to find the joint density of (Y, W). In the Γ-OUcase, the BDLP Z_t is a Compound Poisson process given by:

$$Z_t = \sum_{i=1}^{N_t} b_i \qquad \text{where} \qquad b_i \sim \text{Gamma}(1, \alpha)$$

and N_t is a Poisson(νt) random variable. This implies we can rewrite Y and W in the following way: conditioning on $N_h = n$, let $0 \leq T_1 < T_2 < \ldots < T_n \leq h$ denote the ordered jump times of Z_t and let $R_i = \Delta Z_{T_i}$ denote the jump size at time T_i. Then

$$Y = \sum_{T_i} (1 - e^{-h+T_i}) R_i \qquad \text{and} \qquad W = \sum_{T_i} R_i \tag{B.13}$$

By this definition,

$$Y \leq (1 - e^{-h}) W \quad \text{a.s.}$$

We will use the joint density of $(T_i,\ R_i)$'s to find the joint density of (Y, W). Conditioning on $N_h = n$, T_i's are distributed as the order statistics of a sample of n Uniform$(0, h)$ random variables. So the joint density function $h_{T,n}$ of T_i's is given by:

$$h_{T,n}(t_1, t_2, \ldots, t_n) \;=\; n!\, \Pi_{i=1}^{n} \frac{1}{h} \mathbb{I}_{\{t_1 < t_2 < \cdots < t_n \leq h\}} \;=\; \frac{n!}{h^n} \mathbb{I}_{\{t_1 < t_2 < \cdots < t_n \leq h\}}$$

Since R_i's are independent (with or without the conditioning) of the t_i's and the variables R_i are jointly independent, the joint density $d_{R,n}$ of $R_1, R_2, \ldots, R_n$ is given by:

$$d_{R,n}(r_1, r_2, \ldots, r_n) \;=\; \Pi_{i=1}^{n} \alpha e^{-\alpha r_i} \mathbb{I}_{\{r_i \geq 0\}}.$$

Therefore, the joint density $f_{T,R,n}$ of $(T_i,\, R_i)$'s is given by:

$$f_{T,R,n}(t_1, \ldots, t_n, r_1, \ldots, r_n) \;=\; \frac{n!}{h^n}\mathbb{I}_{\{t_1 < t_2 < \cdots < t_n \le h\}} \cdot \Pi_{i=1}^{n} \alpha e^{-\alpha r_i}\, \mathbb{I}_{\{r_i \ge 0\}}.$$

Motivated by (B.13), consider the following transform $\mathcal{H}$ from $(T_1, \ldots, T_n, R_1, \ldots, R_n)$ to $(U_1, \ldots, U_n, Y, W, V_3, \ldots, V_n)$:

$$\begin{aligned}
U_i &= T_i \quad \text{for} \quad i = 1, \ldots, n \\
Y &= (1 - e^{-h+T_1})R_1 + (1 - e^{-h+T_2})R_2 + \ldots + (1 - e^{-h+T_n})R_n \\
W &= R_1 + R_2 + \ldots + R_n \\
V_i &= R_i \quad \text{for} \quad i = 3, \ldots, n
\end{aligned}$$

Then its inverse transform $\mathcal{H}^{-1}$ is given by:

$$\begin{aligned}
R_1 = {} & \frac{1}{e^{-h+U_2} - e^{-h+U_1}}\big[Y - (1 - e^{-h+U_2})W\big] \\
& + \frac{1}{e^{-h+U_2} - e^{-h+U_1}}\Big[(1 - e^{-h+U_2})\big[V_3 + \ldots + V_n\big] \\
& \qquad - \big[(1 - e^{-h+U_3})V_3 + \ldots + (1 - e^{-h+U_n})V_n\big]\Big] \\
R_2 = {} & \frac{1}{e^{-h+U_2} - e^{-h+U_1}}\big[(1 - e^{-h+U_1})W - Y\big] \\
& + \frac{1}{e^{-h+U_2} - e^{-h+U_1}}\Big[\big[(1 - e^{-h+U_3})V_3 + \ldots + (1 - e^{-h+U_n})V_n\big] \\
& \qquad - (1 - e^{-h+U_1})\big[V_3 + \ldots + V_n\big]\Big]
\end{aligned}$$

and

$$\begin{aligned}
T_i &= U_i \quad \text{for} \quad i = 1, \ldots, n \\
R_i &= V_i \quad \text{for} \quad i = 3, \ldots, n
\end{aligned}$$

which can be written in a more compact form:

$$
\begin{aligned}
&T_i = U_i \quad \text{for} \quad i = 1, \ldots, n \\
&R_i = V_i \quad \text{for} \quad i = 3, \ldots, n \\
&R_1 = \frac{1}{e^{-h+U_2} - e^{-h+U_1}} \Big[Y - (1 - e^{-h+U_2}) W + \sum_{i=3}^{n} (e^{-h+U_i} - e^{-h+U_2}) V_i \Big] \\
&R_2 = \frac{1}{e^{-h+U_2} - e^{-h+U_1}} \Big[-Y + (1 - e^{-h+U_1}) W + \sum_{i=3}^{n} (e^{-h+U_1} - e^{-h+U_i}) V_i \Big]
\end{aligned}
$$

Recall R_1 and R_2 are Gamma$(1, \alpha)$ random variables so they are both positive, which implies, for given positive (Y, W), that the V_i's and the ordered U_i's are constrained in the following region $\Xi_n(y, w)$ for $n \geq 3$ and $0 \leq y \leq (1 - e^{-h})w$:

$$
\begin{aligned}
\Xi_n(y, w) \triangleq \Big\{ (v_3, \ldots, v_n, u_1, \ldots, u_n) :\ & 0 \leq u_1 < u_2 < \cdots < u_n \leq h, \\
& y - (1 - e^{-h+u_2}) w + \sum_{i=3}^{n} (e^{-h+u_i} - e^{-h+u_2}) v_i \geq 0 \\
& \text{and} \quad -y + (1 - e^{-h+u_1}) w + \sum_{i=3}^{n} (e^{-h+u_1} - e^{-h+u_i}) v_i \geq 0 \Big\}
\end{aligned}
$$

that is

$$
\begin{aligned}
\Xi_n(y, w) = \Big\{ (v_3, \ldots, v_n, u_1, \ldots, u_n) :\ & 0 \leq u_1 < u_2 < \cdots < u_n \leq h, \\
& \sum_{i=3}^{n} (e^{-h+u_i} - e^{-h+u_2}) v_i \geq (1 - e^{-h+u_2}) w - y \\
& \text{and} \quad \sum_{i=3}^{n} (e^{-h+u_i} - e^{-h+u_1}) v_i \leq (1 - e^{-h+u_1}) w - y \Big\}
\end{aligned}
\tag{B.14}
$$

Next compute the Jacobian matrix for $\mathcal{H}^{-1}$.

$$J = \frac{\partial(T_1, \ldots, T_n, R_1, \ldots, R_n)}{\partial(U_1, \ldots, U_n, Y, W, V_3, \ldots, V_n)}$$

$$= \begin{array}{c} \\ T_1 \\ T_2 \\ \vdots \\ T_n \\ R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_n \end{array} \begin{array}{c} \begin{array}{ccccccccc} U_1 & U_2 & \ldots & U_n & Y & W & V_3 & \ldots & V_n \end{array} \\ \begin{pmatrix} 1 & & & & & & & & \\ & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & 1 & & & & & \\ * & * & * & * & J_{1Y} & J_{1W} & * & * & * \\ * & * & * & * & J_{2Y} & J_{2W} & * & * & * \\ & & & & & & 1 & & \\ & & & & & & & \ddots & \\ & & & & & & & & 1 \end{pmatrix} \end{array}$$

The empty elements in the matrix should be understood as 0, and the $*$'s represent some non-trivial derivatives which do not contribute to the determinant of J. One can show

$$J_{iY} = \frac{\partial R_i}{\partial Y} = (-1)^{i-1} \frac{1}{e^{-h+U_2} - e^{-h+U_1}}$$

and

$$J_{1W} = \frac{\partial R_1}{\partial W} = -\frac{1 - e^{-h+U_2}}{e^{-h+U_2} - e^{-h+U_1}}, \qquad J_{2W} = \frac{\partial R_2}{\partial W} = \frac{1 - e^{-h+U_1}}{e^{-h+U_2} - e^{-h+U_1}}$$

So the determinant of J can be computed:

$$\begin{aligned}
|J| &= J_{1Y} \cdot J_{2W} - J_{2Y} \cdot J_{1W} \\
&= \frac{1 - e^{-h+U_1}}{(e^{-h+U_2} - e^{-h+U_1})^2} - \frac{1 - e^{-h+U_2}}{(e^{-h+U_2} - e^{-h+U_1})^2} \\
&= \frac{e^{-h+U_2} - e^{-h+U_1}}{(e^{-h+U_2} - e^{-h+U_1})^2} \\
&= \frac{1}{e^{-h+U_2} - e^{-h+U_1}}
\end{aligned}$$

Therefore, the joint density of $(U_1, \ldots, U_n, Y, W, V_3, \ldots, V_n)$ conditionally given $N_h = n$ is:

$$\begin{aligned}
&f_{(Y,W,U_1,\ldots,U_n,V_3,\ldots,V_n)|N_h=n}(y, w, u_1, \ldots, u_n, v_3, v_n) \\
&= \frac{n!}{h^n}\ \mathbb{I}_{\{(\mathbf{v},\mathbf{u}) \in \Xi_n(y,w)\}} \\
&\quad \cdot \alpha\ \exp\Big\{ - \frac{\alpha}{e^{-h+u_2} - e^{-h+u_1}} \big[y - (1 - e^{-h+u_2})w + \sum_{i=3}^{n} (e^{-h+u_i} - e^{-h+u_2}) v_i \big] \Big\} \\
&\quad \cdot \alpha\ \exp\Big\{ - \frac{\alpha}{e^{-h+u_2} - e^{-h+u_1}} \big[-y + (1 - e^{-h+u_1})w + \sum_{i=3}^{n} (e^{-h+u_1} - e^{-h+u_i}) v_i \big] \Big\} \\
&\quad \cdot \alpha^{n-2}\ \exp\Big\{ -\alpha \sum_{i=3}^{n} v_i \Big\}.
\end{aligned}$$

Simplifying the expression,

$$\begin{aligned}
&f_{(Y,W,\mathbf{U},\mathbf{V})|N_h=n}(y, w, u_1, \ldots, u_n, v_3, v_n) \\
&= \alpha^2\ \exp\Big\{ -\alpha w - \frac{\alpha}{e^{-h+u_2} - e^{-h+u_1}} \sum_{i=3}^{n} \big((e^{-h+u_i} - e^{-h+u_2}) + (e^{-h+u_1} - e^{-h+u_i}) \big) v_i \Big\} \\
&\quad \cdot \alpha^{n-2}\ \exp\Big\{ -\alpha \sum_{i=3}^{n} v_i \Big\} \cdot \frac{n!}{h^n}\ \mathbb{I}_{\{(\mathbf{v},\mathbf{u}) \in \Xi_n(y,w)\}} \\
&= \frac{n!}{h^n} \cdot \alpha^n e^{-\alpha w}\ \mathbb{I}_{\{(\mathbf{v},\mathbf{u}) \in \Xi_n(y,w)\}} \qquad \text{(B.15)}
\end{aligned}$$

The joint density of $(Y, W)|N_h = n$ is given by:

$$f_{(Y,W)|N_h=n}(y, w) = \frac{n!}{h^n} \cdot \alpha^n e^{-\alpha w} \iint_{\Xi_n(y,w)} \mathrm{d}\mathbf{v}\, \mathrm{d}\mathbf{u}$$

In the case when $N_h = 1$ and $N_h = 2$, the joint density of $(Y, W)|N_h$ has a slight different form. We derive them separately next.

- In the case $N_h = 1$, let T_1 denote the unique jump time in $[0, h]$, and let $R_1 = \Delta Z_{T_1}$. Then the transform from (R_1, T_1) to (Y, W) is given by:

$$W = R_1, \qquad Y = (1 - e^{-h+T_1})R_1$$

with the inverse transform and the Jacobian J:

$$R_1 = W, \qquad T_1 = h + \ln(1 - \frac{Y}{W}), \qquad J = \frac{1}{Y - W}.$$

Using the fact that the joint density of (R_1, T_1) is

$$f_{(R_1,T_1)}(r_1, t_1) = \alpha\, e^{-\alpha r_1} \frac{1}{h}\, \mathbb{I}_{\{0 \le t_1 \le h\}},$$

the joint density of (Y, W) is given by

$$\begin{aligned} f_{(Y,W)|N_h=1}(y, w) &= \frac{\alpha}{h}\, e^{-\alpha w} \frac{1}{w - y}\, \mathbb{I}_{\{0 \le h + \ln(1 - y/w) \le h\}} \\ &= \frac{\alpha}{h}\, e^{-\alpha w} \frac{1}{w - y}\, \mathbb{I}_{\{0 \le y \le (1-e^{-h})w\}}. \end{aligned}$$

- Now, in the case $N_h = 2$, let T_1 and T_2 denote the ordered jump times in $[0, h]$ and let $R_i = \Delta Z_{T_i}$ for $i = 1, 2$. Using a change variables similar to that used when $N_h \geq 3$, one can derive the conditional joint density of (Y, W, U_1, U_2) in the form:

$$f_{(Y,W,U_1,U_2)|N_h=2}(y, w, u_1, u_2) = \alpha^2\, e^{-\alpha w}\, \frac{2}{h^2}\, \mathbb{I}_{\{(u_1,u_2)\in\Xi_2(y,w)\}}$$

where $\Xi_2(y, w)$ is defined to be the region where

$$\Xi_2(y, w) = \Big\{(u_1, u_2) :\ 0 \leq u_1 \leq h \,+\, \ln(1 - \frac{y}{w}) \\ \text{and } h \,+\, \ln(1 - \frac{y}{w}) \leq u_2 \leq h\Big\}$$

Therefore, the joint density of (Y, W) is given by:

$$f_{(Y,W)|N_h=2}(y, w) = \frac{2\alpha^2}{h^2}\, e^{-\alpha w}\, \ln\Big(1 - \frac{y}{w}\Big)^{-1} \Big(h + \ln\Big(1 - \frac{y}{w}\Big)\Big)\, \mathbb{I}_{\{0\leq y\leq(1-e^{-h})w\}}$$

Then the joint density of (Y, W) can be derived by unconditioning on N_h:

$$\begin{aligned} f_{Y,W}(y, w) &= \sum_{n=0}^{\infty} f_{(Y,W)|N_h=n}(y, w) \cdot \frac{e^{-\nu}\nu^n}{n!} \\ &= e^{-\alpha w}\, \frac{\alpha\, e^{-\nu}}{h}\, \frac{1}{w - y}\, \mathbb{I}_{\{0\leq y\leq(1-e^{-h})w\}} \\ &\quad + e^{-\alpha w}\, \frac{\alpha^2\, e^{-\nu}}{h^2}\, \ln\Big(1 - \frac{y}{w}\Big)^{-1} \Big(h + \ln\Big(1 - \frac{y}{w}\Big)\Big)\, \mathbb{I}_{\{0\leq y\leq(1-e^{-h})w\}} \\ &\quad + e^{-\alpha w} e^{-\nu} \sum_{n=3}^{\infty} \Big(\frac{\alpha\nu}{h}\Big)^n \int_{\Xi_n(y,w)} \mathrm{d}\mathbf{v}\, \mathrm{d}\mathbf{u}. \end{aligned} \tag{B.16}$$

Finally, we have the joint density of (S, Y, W):

$$f_{Y,W,S}(y,w,s) = \frac{\alpha^\nu}{\Gamma(\nu)} s^{\nu-1} e^{-\alpha s} f_{Y,W}(y,w) \tag{B.17}$$

The joint density of X_1 under the Gamma BN-S model is given by the following integral expression:

$$m_{\boldsymbol{\theta}}(x) = \int_0^\infty \int_0^\infty \int_0^{(1-e^{-h})w} \frac{f_{Y,W,S}(y,w,s)}{\sqrt{2\pi}\sqrt{(1-e^{-h})s+y}} \exp\left\{-\frac{\left(x-\mu-\beta((1-e^{-h})s+y)-\rho w\right)^2}{2((1-e^{-h})s+y)}\right\} \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \tag{B.18}$$

To simplify the notations, let $g(x, y, w, s; \boldsymbol{\theta})$ and $\mathbb{D}$ denote the integrand and the integration region over (y, w, s) in (B.18) respectively. To study the derivatives of $m_{\boldsymbol{\theta}}(x)$ with respect to $\boldsymbol{\theta}$, it is equivalent to study the derivatives of $g(x, y, w, s; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$. Introduce the following notations:

$$\partial_1 = \frac{\partial}{\partial\mu}, \quad \partial_2 = \frac{\partial}{\partial\beta} \quad \partial_3 = \frac{\partial}{\partial\rho}, \quad \partial_4 = \frac{\partial}{\partial\nu} \quad \text{and} \quad \partial_5 = \frac{\partial}{\partial\alpha}$$

Notations for higher order partial (cross) derivatives are defined similarly. For example, $\partial_2^{p_2} = \frac{\partial^{p_2}}{\partial\alpha^{p_2}}$ and $\partial_{35} = \frac{\partial^2}{\partial\rho\partial\alpha}$. The following proposition summarizes a useful result.

Proposition B.2.1: All partial derivatives and cross derivatives of $g(x, y, w, s; \boldsymbol{\theta})$

with respect to $\boldsymbol{\theta}$ of arbitrary order p can be written in the following form:

$$\frac{\partial^p \, g(x,y,w,s;\boldsymbol{\theta})}{\partial_1^{p_1} \cdots \partial_5^{p_5}} = \sum_{\mathbf{a}} \Psi_{1,\mathbf{a}}(h,\boldsymbol{\theta}) \, g(x,y,w,s;\boldsymbol{\theta}) \, \frac{x^{a_1} \, y^{a_2} \, w^{a_3} \, s^{a_4} \, (\ln s)^{a_5}}{((1-e^{-h})s + y))^{a_6}} \tag{B.19}$$

where the summation over $\mathbf{a}$ is finite and

- $\Psi_{1,\mathbf{a}}(h,\boldsymbol{\theta})$ is a generic function of the parameters h and $\boldsymbol{\theta}$ and the subscript $\mathbf{a} = (a_1, a_2, a_3, a_4, a_5, a_6)$;
- p_i's are non-negative integers where $p_1 + p_2 + \ldots + p_5 = p$ for $p \geq 1$.
- a_1 to a_6 are non-negative integers such that a_1 and a_6 are less than $(p_1 + p_2 + p_3)$, $a_2 \leq (p_1 + 2p_2 + p_3)$, $a_3 \leq (p_1 + p_2 + 2p_3)$, $a_4 \leq (p_1 + 2p_2 + p_3 + p_5)$ and $a_5 \leq p_4$.

Proof: To study partial derivatives of $g(x,y,w,s;\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ of arbitrary order p, it is sufficient to compute the first order derivatives and derive the general patterns from them. For partial derivatives with respect to μ, β and ρ,

$$\begin{aligned}
\partial_1 \, g(x,y,w,s;\boldsymbol{\theta}) &= g(x,y,w,s;\boldsymbol{\theta}) \, \frac{x - \mu - \beta((1-e^{-h})s + y) - \rho w}{(1-e^{-h})s + y} \\
\partial_2 \, g(x,y,w,s;\boldsymbol{\theta}) &= g(x,y,w,s;\boldsymbol{\theta}) \, \frac{x - \mu - \beta((1-e^{-h})s + y) - \rho w}{(1-e^{-h})s + y} \cdot ((1-e^{-h})s + y) \\
&= g(x,y,w,s;\boldsymbol{\theta}) \, \big(x - \mu - \beta((1-e^{-h})s + y) - \rho w\big) \\
\partial_3 \, g(x,y,w,s;\boldsymbol{\theta}) &= g(x,y,w,s;\boldsymbol{\theta}) \, \frac{x - \mu - \beta((1-e^{-h})s + y) - \rho w}{(1-e^{-h})s + y} \cdot w
\end{aligned}$$

For partial derivatives with respect to (ν, α), first notice that:

$$\partial_4 f_{Y,W,S}(y, w, s) = \ln\alpha \cdot f_{Y,W,S}(y, w, s) - \frac{\psi_0(\nu)}{\Gamma(\nu)} f_{Y,W,S}(y, w, s)$$
$$+ \ln s \cdot f_{Y,W,S}(y, w, s) - f_{Y,W,S}(y, w, s) + \frac{\alpha}{h} f_{Y,W,S}(y, w, s)$$
$$\partial_5 f_{Y,W,S}(y, w, s) = \frac{\nu}{\alpha}(-s) \cdot f_{Y,W,S}(y, w, s) - w \cdot f_{Y,W,S}(y, w, s) + \frac{\nu}{h} f_{Y,W,S}(y, w, s)$$

where $\psi_0(\nu)$ is the diGamma function. Then,

$$\partial_4 g(x, y, w, s; \boldsymbol{\theta}) = g(x, y, w, s; \boldsymbol{\theta}) \left(\ln s + \ln\alpha + \frac{\alpha}{h} - \frac{\psi_0(\nu)}{\Gamma(\nu)} - 1 \right)$$
$$\partial_5 g(x, y, w, s; \boldsymbol{\theta}) = g(x, y, w, s; \boldsymbol{\theta}) \left(\frac{\nu}{h} - \frac{\nu}{\alpha} s - w \right)$$

From these computations, it is not difficult to see that higher order derivatives have exact expressions obtained by successively differentiating g. For example,

$$\begin{aligned}
&\partial_{225}\, g(x, y, w, s; \boldsymbol{\theta}) \\
&\quad = \partial_5 \Big(g(x, y, w, s; \boldsymbol{\theta}) \left(x - \mu - \beta((1 - e^{-h})s + y) - \rho w\right)^2 \\
&\qquad\qquad - g(x, y, w, s; \boldsymbol{\theta}) \left((1 - e^{-h})s + y\right) \Big) \\
&\quad = g(x, y, w, s; \boldsymbol{\theta}) \left(\frac{\nu}{h} - \frac{\nu}{\alpha} s - w\right) \cdot \left(x - \mu - \beta((1 - e^{-h})s + y) - \rho w\right)^2 \\
&\qquad\qquad - g(x, y, w, s; \boldsymbol{\theta}) \left(\frac{\nu}{h} - \frac{\nu}{\alpha} s - w\right) \cdot \left((1 - e^{-h})s + y\right).
\end{aligned}$$

After expanding the square and cross multipling all terms, the derivative above is in the form of (B.19) with $a_3 = a_6 = 0$, a_1 and a_4 ranging from 1 to 2, a_2 and a_5 ranging from 1 to 3. Further notice that the p-th order derivative of the Gamma function

$$\frac{\mathrm{d}^p}{\mathrm{d}\nu^p}\Gamma(\nu) = \int_0^\infty t^{\nu-1}e^{-t}(\ln t)^p \, \mathrm{d}t$$

is well defined for $\nu > 0$. Therefore, we can conclude that derivatives of $g(x, y, w, s; \boldsymbol{\theta})$ of any order can be written in the form of (B.19). ■

A direct consequence of Proposition B.2.1 is that one can obtain upper bounds for derivatives of $g(x, y, w, s; \boldsymbol{\theta})$. For example,

- $|\,\partial_1\, g(x, y, w, s; \boldsymbol{\theta})| \leq \dfrac{g(x, y, w, s; \boldsymbol{\theta})}{(1-e^{-h})} \left(\dfrac{|x| + |\mu| + |\rho|w}{s} + |\beta|(1-e^{-h})\right)$

- $|\,\partial_2\, g(x, y, w, s; \boldsymbol{\theta})| \leq g(x, y, w, s; \boldsymbol{\theta}) \left(|x| + |\mu| + |\beta|((1-e^{-h})s + y) + |\rho|w\right)$

- $|\,\partial_3\, g(x, y, w, s; \boldsymbol{\theta})| \leq \dfrac{g(x, y, w, s; \boldsymbol{\theta})}{(1-e^{-h})} \left[\dfrac{(|x| + |\mu| + |\rho|w)w}{s} + |\beta|(1-e^{-h})w\right]$

- $|\,\partial_4\, g(x, y, w, s; \boldsymbol{\theta})| \leq g(x, y, w, s; \boldsymbol{\theta}) \left(\big|\ln\alpha\big| + \dfrac{\psi_0(\nu)}{\Gamma(\nu)} + 1 + \dfrac{\alpha}{h} + |\ln s|\right)$

- $|\,\partial_5\, g(x, y, w, s; \boldsymbol{\theta})| \leq g(x, y, w, s; \boldsymbol{\theta}) \left(\dfrac{\nu}{\alpha}s + w + \dfrac{\nu}{h}\right)$

The next proposition summarizes a general result.

Proposition B.2.2: All partial and cross derivatives of $g(x, y, w, s; \boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ can be bounded by functions in the following form:

$$\left|\frac{\partial^p\, g(x, y, w, s; \boldsymbol{\theta})}{\partial_1^{p_1} \cdots \partial_5^{p_5}}\right| \leq \sum_{l,q,r} \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |x|^l\, s^q\, w^r\, g(x, y, w, s; \boldsymbol{\theta}) \tag{B.20}$$

where $\mathrm{i} = (l, q, r)$ is a vector of integers such that $0 \leq l \leq p_1 + p_2 + p_3$, $-(p_1 + p_2 + p_3) \leq q \leq (p_2 + p_4 + p_5)$ and $0 \leq r \leq p_1 + 2(p_2 + p_3)$. Here $\Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})$ is again a generic function of h and $\boldsymbol{\theta}$ and is continuous over $\boldsymbol{\theta}$ for $\boldsymbol{\theta} \in B_a$.

Proof: From the derivation of Proposition B.2.1, one can find the following patterns:

- Differentiating $\theta_1(\mu)$, $\theta_2(\beta)$ and $\theta_3(\rho)$ increases the order of $|x|$ and $\frac{1}{(1-e^{-h})s+y}$ by 1 respectively. But since the term $\frac{1}{(1-e^{-h})s+y}$ is bounded above by $\frac{1}{(1-e^{-h})s}$, we will focus on the order change of $\frac{1}{s}$ only.

- Differentiating $\theta_2(\beta)$ increases the order of s and y by 2.

- Differentiating $\theta_3(\rho)$ increases the order of w by 2.

- Differentiating $\theta_4(\nu)$ increases the order of $\ln s$ by 1. But as $|\ln s|$ is bounded by $\frac{1}{s}\,\mathbb{I}_{\{s\leq 1\}} + s\,\mathbb{I}_{\{s>1\}}$, we can treat the effect of this differentiation as increasing the order of s and $\frac{1}{s}$ by 1 respectively.

- Differentiating $\theta_5(\alpha)$ increases the order of s by 1.

Now, by (B.19),

$$
\begin{aligned}
\left|\frac{\partial^p\, g(x,y,w,s;\boldsymbol{\theta})}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\right| &\le \sum_{\mathbf{a}} \Psi_{1,\mathbf{a}}(h,\boldsymbol{\theta})\, g(x,y,w,s;\boldsymbol{\theta})\, \frac{|x|^{a_1}\, s^{a_2}\, |\ln s|^{a_3}\, y^{a_4}\, w^{a_5}}{((1-e^{-h})s+y))^{a_6}} \\
&\le \sum_{\mathbf{a}} \frac{\Psi_{1,\mathbf{a}}(h,\boldsymbol{\theta})}{(1-e^{-h})^{a_6}}\, g(x,y,w,s;\boldsymbol{\theta})\, \frac{|x|^{a_1}\, s^{a_2}\, |\ln s|^{a_3}\, y^{a_4}\, w^{a_5}}{s^{a_6}} \\
&\le \sum_{\mathbf{a}} \frac{\Psi_{1,\mathbf{a}}(h,\boldsymbol{\theta})}{(1-e^{-h})^{a_6-a_4}}\, g(x,y,w,s;\boldsymbol{\theta})\, \left(|x|^{a_1}\, s^{a_2-a_6}\, |\ln s|^{a_3}\, w^{a_4+a_5}\right) \\
&\le \frac{\Psi_{1,\mathbf{a}}(h,\boldsymbol{\theta})\, a_3!}{(1-e^{-h})^{a_6-a_4}}\, g(x,y,w,s;\boldsymbol{\theta}) \\
&\qquad \cdot \sum_{\mathbf{a}} \left(|x|^{a_1}\, s^{a_2+a_3-a_6}\, w^{a_4+a_5} + |x|^{a_1}\, s^{a_2-a_3-a_6}\, w^{a_4+a_5}\right)
\end{aligned}
$$

The third inequality holds since $y \le (1-e^{-h})w$ and $(s+\frac{1}{s})^q \le q!(s^q + \frac{1}{s^q})$. Therefore, combining the patterns described above for the successive differentiations and converting $\mathbf{a}$ to index $\mathbf{i} = (l,q,r)$, one finds (B.20) holds with the given range on l, q and r. ■

Remark In fact, the upper bounds of the ranges of the indices are not critical as all positive moments of S and W are finite. However, the minimum value which q can take is important, because the negative q-th moment of a Gamma(ν,α) random variable is finite only when $\nu > q$.

Next, we will use the following lemma to find the derivatives of $m_{\boldsymbol{\theta}}(x)$ with respect to $\boldsymbol{\theta}$. This lemma will be used throughout the rest of this section to justify the validity of interchanging integration and differentiation.

Lemma B.2.3: **(Billingsley 1995, Theorem 16.8)** Let Θ be an open subset of $\mathbb{R}$ and $\mathcal{S}$ be a measure space. Suppose that a function $f : \Theta \times \mathcal{S} \to \mathbb{R}$ satisfies the following conditions:

(i) $f(\vartheta, x)$ is a measurable function of ϑ and x jointly, and is integrable over x, for almost all $\vartheta \in \Theta$ held fixed.

(ii) For almost all $x \in \mathcal{S}$, $f(\vartheta, x)$ is an absolutely continuous function of ϑ.

(iii) $\partial f / \partial \vartheta$ is "locally integrable", that is, for all compact intervals $[a, b] \in \Theta$:

$$\int_a^b \int_{\mathcal{S}} \left| \frac{\partial}{\partial \vartheta} f(\vartheta, x) \right| \mathrm{d}x \ \mathrm{d}\vartheta < \infty \tag{B.21}$$

Then $\int_{\mathcal{S}} f(\vartheta, x) \, \mathrm{d}x$ is an absolutely continuous function of ϑ, and for almost every $\vartheta \in \Theta$, its derivative exists and is given by

$$\frac{\partial}{\partial \vartheta} \int_{\mathcal{S}} f(\vartheta, x) \, \mathrm{d}x = \int_{\mathcal{S}} \frac{\partial}{\partial \vartheta} f(\vartheta, x) \, \mathrm{d}x$$

In regard to the partial derivatives of $m_{\boldsymbol{\theta}}(x)$ with respect to $\boldsymbol{\theta}$, the following result holds.

Proposition B.2.4: If $\nu > 7/2$, then the partial (cross) derivatives for $m_{\boldsymbol{\theta}}(x)$ of order up to 3 can be computed by interchanging the differentiation and integration in expression (B.18), i.e.

$$\frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}} m_{\boldsymbol{\theta}}(x) = \iiint_{\mathbb{D}} \frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}} g(x, y, w, s; \boldsymbol{\theta}) \ \mathrm{d}y \ \mathrm{d}w \ \mathrm{d}s$$

Proof: Observe that:

- $g(x, y, w, s; \boldsymbol{\theta})$ is measurable in both (y, w, s) and $\boldsymbol{\theta}$, and it is integrable over (y, w, s) with $\boldsymbol{\theta}$ held fixed when $\nu > 0$.

- From Proposition B.2.2, $\left|\dfrac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}}\ g(x, y, w, s; \boldsymbol{\theta})\right|$ is bounded over $\boldsymbol{\theta}$ for finite (x, y, w, s) when $\boldsymbol{\theta} \in B_a$, so $g(x, y, w, s; \boldsymbol{\theta})$ is an absolutely continuous function of $\boldsymbol{\theta}$.

- Recall B_a is the 5-dimensional sphere centered at $\boldsymbol{\theta}_s$ with radius a. Let $\int_{B,a}$ denote the integration of $\boldsymbol{\theta} = (\mu, \beta, \rho, \nu, \alpha)$ in the sphere B_a. By (B.20),

$$\begin{aligned}
&\int_{B_a} \iiint_{\mathbb{D}} \left|\frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}}\ g(x, y, w, s; \boldsymbol{\theta})\right| \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s\ \mathrm{d}\boldsymbol{\theta} \\
\leq &\int_{B_a} \iiint_{\mathbb{D}} \sum_{l,q,r} \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |x|^l\, s^q\, w^r\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s\ \mathrm{d}\boldsymbol{\theta} \qquad (\triangle)
\end{aligned}$$

Recall the exponential term in $g(x, y, w, s; \boldsymbol{\theta})$ (B.18) is bounded by 1, and

$$\frac{1}{\sqrt{(1-e^{-h})s + y}} \leq \frac{1}{\sqrt{(1-e^{-h})s}}.$$

$$\begin{aligned}
\triangle &< \sum_{l,q,r} \int_{B_a} \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |x|^l \iiint_{\mathbb{D}} s^q\, w^r \frac{f_{Y,W,S}(y, w, s)}{\sqrt{(1-e^{-h})s + y)}}\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s\ \mathrm{d}\boldsymbol{\theta} \\
&< \sum_{l,q,r} \int_{B_a} \frac{\Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |x|^l}{\sqrt{1-e^{-h}}} \iiint_{\mathbb{D}} s^{q-1/2}\, w^r\, f_{Y,W,S}(y, w, s)\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s\ \mathrm{d}\boldsymbol{\theta} \\
&= \sum_{l,q,r} \frac{|x|^l}{\sqrt{1-e^{-h}}} \int_{B_a} \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, E[S^{q-1/2} W^r]\, \mathrm{d}\boldsymbol{\theta} \\
&< \infty
\end{aligned}$$

Since $S \sim \mathrm{Gamma}(\nu, \alpha)$ with $\nu > 7/2$ and $q \geq -3$ by Proposition B.2.2, $E[S^{q-1/2}]$ is finite. Further, W is a Compound Poisson random variable with all positive moments $E[W^r]$ finite, therefore the last inequality holds.

According to Lemma (B.2.3), we can conclude that

$$\frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, m_{\boldsymbol{\theta}}(x) = \iiint_{\mathbb{D}} \frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, g(x, y, w, s; \boldsymbol{\theta})\ \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s$$

for $p \leq 3$ when $\nu > 7/2$. ■

The result in Proposition B.2.4 can be extended to study the derivatives of $m^*_{\boldsymbol{\theta}}(x)$.

Proposition B.2.5: If $\nu > 7/2$, then the partial (cross) derivatives for $m^*_{\boldsymbol{\theta}}(x)$ of order up to 3 can be computed by interchanging the differentiation and integration, i.e.

$$\frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, m^*_{\boldsymbol{\theta}}(x) = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-t)^2}{2}} \cdot \frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, m_{\boldsymbol{\theta}}(t)\ \mathrm{d}t$$

Proof: It is easy to see that function $e^{-\frac{(x-t)^2}{2}} m_{\boldsymbol{\theta}}(t)$ is a measurable function for both t and $\boldsymbol{\theta}$ and integrable over t when $\boldsymbol{\theta}$ held fixed. Using the proof of Proposition B.2.4,

$$\begin{aligned}
&\left|\frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, e^{-\frac{(x-t)^2}{2}} m_{\boldsymbol{\theta}}(t)\right| \leq \left|\frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, m_{\boldsymbol{\theta}}(t)\right| \\
&\leq \iiint_{\mathbb{D}} \left|\frac{\partial^p}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\, g(t, y, w, s; \boldsymbol{\theta})\right| \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&< \sum_{l,q,r} \frac{\Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |t|^l}{\sqrt{1-e^{-h}}} \iiint_{\mathbb{D}} s^{q-1/2}\, w^r\, f_{Y,W,S}(y, w, s)\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= \sum_{l,q,r} \frac{|t|^l}{\sqrt{1-e^{-h}}}\, \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, E[S^{q-1/2} W^r] < \infty
\end{aligned}$$

One concludes that $e^{-\frac{(x-t)^2}{2}} m_{\boldsymbol{\theta}}(t)$ is absolutely continuous with respect to $\boldsymbol{\theta}$ for all finite t. To verify condition (B.21), notice

$$
\begin{aligned}
&\int_{B_a}\int_{\mathbb{R}}\left|\frac{\partial^p\, e^{-\frac{(x-t)^2}{2}} m_{\boldsymbol{\theta}}(t)}{\partial_1^{p_1}\cdots\partial_5^{p_5}}\right| \mathrm{d}t\ \mathrm{d}\boldsymbol{\theta} \\
&< \int_{B_a}\int_{\mathbb{R}} e^{-\frac{(x-t)^2}{2}} \sum_{l,q,r} \frac{|t|^l}{\sqrt{1-e^{-h}}}\, \Psi_{2,\mathrm{i}}(h,\boldsymbol{\theta})\, E[S^{q-1/2}W^r]\, \mathrm{d}t\ \mathrm{d}\boldsymbol{\theta} \\
&= \sum_{l,q,r}\int_{\mathbb{R}} e^{-\frac{(x-t)^2}{2}} \frac{|t|^l}{\sqrt{1-e^{-h}}}\, \mathrm{d}t \int_{B_a} \Psi_{2,\mathrm{i}}(h,\boldsymbol{\theta})\, E[S^{q-1/2}W^r]\, \mathrm{d}\boldsymbol{\theta} \\
&< \infty
\end{aligned}
$$

for all finite x. Therefore, the conditions in Lemma B.2.3 are satisfied and the result in the Proposition holds. ■.

Knowing how to compute the derivatives of $m^*_{\boldsymbol{\theta}}(x)$, we turn to examine the Taylor expansion in Lemma 3.2.3. The integrand of the disparity, i.e., $G(\delta^n)m^*_{\boldsymbol{\theta}}(x)$, is a measurable function of $\boldsymbol{\theta}$ and x jointly, and it is integrable over x when $\boldsymbol{\theta}$ held fixed. Therefore, to study the interchange of differentiation and integration, one needs to verify the following results when $\boldsymbol{\theta}$ is in the neighborhood of $\boldsymbol{\theta}^s$:

(I1). The density $m^*_{\boldsymbol{\theta}}(x)$ is absolutely continuous with respect to θ_i, (θ_i,θ_j) and $(\theta_i,\theta_j,\theta_k)$ for all i, j and k ranging from 1 to 5.

(I2). Condition (B.21) holds with $f(\vartheta,x)$ replaced by functions of the following forms

- $\partial_i m^*_{\boldsymbol{\theta}}(x)$, $\quad \partial_{ij} m^*_{\boldsymbol{\theta}}(x)$, $\quad \partial_{ijk} m^*_{\boldsymbol{\theta}}(x)$
- $\partial_{ij} m^*_{\boldsymbol{\theta}}(x)\cdot\partial_k m^*_{\boldsymbol{\theta}}(x)$
- $\partial_i \ln m^*_{\boldsymbol{\theta}}(x)\cdot\partial_j \ln m^*_{\boldsymbol{\theta}}(x)\cdot m^*_{\boldsymbol{\theta}}(x)$

- $\partial_i \ln m^*_{\boldsymbol{\theta}}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}}(x) \cdot \partial_k \ln m^*_{\boldsymbol{\theta}}(x) \cdot m^*_{\boldsymbol{\theta}}(x)$

- $\partial_{ij} m^*_{\boldsymbol{\theta}}(x) \cdot \partial_k \ln m^*_{\boldsymbol{\theta}}(x) \cdot m^*_{\boldsymbol{\theta}}(x)$

Let us motivate the idea of the derivation by looking at the expression (3.11) in Lemma 3.2.3:

$$\begin{aligned}\partial_{ij}\rho(f^*, m^*_{\theta^s}) = &\int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\theta^s}}+1} \cdot \frac{f^*(x)^2}{m^*_{\theta^s}(x)^2} \cdot \partial_i \ln m^*_{\theta^s}(x) \cdot \partial_j \ln m^*_{\theta^s}(x) \cdot m^*_{\theta^s}(x)\, \mathrm{d}x \\ &+ \int_{\mathbb{R}} e^{-\frac{f^*}{m^*_{\theta^s}}+1} \cdot \frac{f^*(x)}{m^*_{\theta^s}(x)} \cdot \partial_{ij} m^*_{\theta^s}(x)\, \mathrm{d}x \\ &+ \int_{\mathbb{R}} (e^{-\frac{f^*}{m^*_{\theta^s}}+1} - 2)\, \partial_{ij} m^*_{\theta^s}(x)\, \mathrm{d}x\end{aligned}$$

In order to derive (3.11) by interchanging the differentiation and integration, one needs to show (I1) and (I2) hold for i, j ranging from 1 to 5:

- For (I1) to hold, $m^*_{\boldsymbol{\theta}}(x)$ needs to be absolutely continuous with respect to θ_i and (θ_i, θ_j). One can show $|\,\partial_i m^*_{\boldsymbol{\theta}}(x)|$, $|\,\partial_j m^*_{\boldsymbol{\theta}}(x)|$ and $|\,\partial_{ij} m^*_{\boldsymbol{\theta}^s}(x)|$ are bounded by some continuous functions of $\boldsymbol{\theta}$, i.e.

$$|\,\partial_i m^*_{\boldsymbol{\theta}}(x)| \leq K_i(\boldsymbol{\theta}), \quad |\,\partial_j m^*_{\boldsymbol{\theta}}(x)| \leq K_j(\boldsymbol{\theta}) \quad \text{and} \quad |\,\partial_{ij} m^*_{\boldsymbol{\theta}}(x)| \leq K_{ij}(\boldsymbol{\theta}).$$

- For (I2) to hold, only the integrals

$$\int_{B_a}\int_{\mathbb{R}} \left|\,\partial_i \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot \partial_j \ln m^*_{\boldsymbol{\theta}^s}(x) \cdot m^*_{\boldsymbol{\theta}^s}(x)\right| \mathrm{d}x\ \mathrm{d}\theta_i\, \mathrm{d}\theta_j$$

and

$$\int_{B_a}\int_{\mathbb{R}} \left|\,\partial_{ij} m^*_{\boldsymbol{\theta}^s}(x)\right| \mathrm{d}x\ \mathrm{d}\theta_i\, \mathrm{d}\theta_j$$

are required to be finite. If one can establish that,

$$\max\big(|\,\partial_i \ln m_{\boldsymbol{\theta}}^*(x)|, |\,\partial_j \ln m_{\boldsymbol{\theta}}^*(x)|\big) \;\le\; C(1+|x|^l)$$

and

$$|\,\partial_{ij} \ln m_{\boldsymbol{\theta}}^*(x)| \;\le\; C_{ij}(1+|x|^{l_{ij}}),$$

for some large constants C and C_{ij} and positive integers l and l_{ij}, and if further $E^*[|X|^l]$, $E^*[|X|^{2l}]$ and $E^*[|X|^{l_{ij}}]$ are finite and continuous functions of (θ_i, θ_j) in the neighborhood of $\boldsymbol{\theta}^s$, then

$$\begin{aligned}\int_{B_a}\int_{\mathbb{R}} \Big|\,\partial_{ij} \ln m_{\boldsymbol{\theta}}^*(x)\Big| \;\mathrm{d}x\;\mathrm{d}\theta_i\,\mathrm{d}\theta_j &= \int_{B_a}\int_{\mathbb{R}} \Big|\,\partial_{ij} m_{\boldsymbol{\theta}}^*(x)\Big| \cdot m_{\boldsymbol{\theta}}^*(x)\,\mathrm{d}x\;\mathrm{d}\theta_i\,\mathrm{d}\theta_j \\ &\le \int_{B_a}\int C_{ij}\;(1+|x|^{l_{ij}})\cdot m_{\boldsymbol{\theta}}^*(x)\,\mathrm{d}x\;\mathrm{d}\theta_i\,\mathrm{d}\theta_j \\ &= \int_{B_a} C_{ij}\,\big(1+E^*[|X|^{l_{ij}}]\big)\,\mathrm{d}\theta_i\,\mathrm{d}\theta_j \\ &< \infty\end{aligned}$$

and

$$\begin{aligned}&\int_{B_a}\int_{\mathbb{R}} \Big|\,\partial_i \ln m_{\boldsymbol{\theta}}^*(x)\cdot\,\partial_j \ln m_{\boldsymbol{\theta}}^*(x)\Big| \cdot m_{\boldsymbol{\theta}}^*(x)\,\mathrm{d}x\;\mathrm{d}\theta_i\,\mathrm{d}\theta_j \\ &\le \int_{B_a}\int C^2\,(1+|x|^l)^2\cdot m_{\boldsymbol{\theta}}^*(x)\,\mathrm{d}x\;\mathrm{d}\theta_i\,\mathrm{d}\theta_j \\ &= \int_{B_a} C^2\,\big(1+2E^*[|X|^l]+E^*[|X|^{2l}]\big)\,\mathrm{d}\theta_i\,\mathrm{d}\theta_j \\ &< \infty\end{aligned}$$

Here, $E^*[\cdot]$ refers to expectation taken with respect to $m_{\boldsymbol{\theta}}^*(x)$. Then the local integrability condition (B.21) over (θ_i, θ_j) can be established and Lemma B.2.3

applies. $\square$

There are two key components in the derivation. The first one is to show the boundedness of derivatives of $m^*_{\boldsymbol{\theta}}(x)$, the other one is to analyze the tail behavior of $\partial_i \ln m^*_{\boldsymbol{\theta}}(x)$ for large $|x|$. Since we use the Gaussian kernel, the following result shows the equivalence between $m^*_{\boldsymbol{\theta}}(x)$ and $m_{\boldsymbol{\theta}}(x)$ in verifying these two components.

Proposition B.2.6: Considering the model density $m_{\boldsymbol{\theta}}(x)$ and the smoothed density $m^*_{\boldsymbol{\theta}}(x)$ in the Γ-OU BN-S model, if $\nu > 7/2$ and $\boldsymbol{\theta}$ is in a compact subset of the sample space, then the following results hold.

(e1). If $|\,\partial_i m_{\boldsymbol{\theta}}(x)|$ is integrable with respect to x, so is $|\,\partial_i m^*_{\boldsymbol{\theta}}(x)|$.

(e2). If $|\,\partial_i m_{\boldsymbol{\theta}}(x)|$ is bounded by a continuous function $K_i(\boldsymbol{\theta})$, then $|\,\partial_i m^*_{\boldsymbol{\theta}}(x)|$ is also bounded by $K_i(\boldsymbol{\theta})$.

(e3). If further $|\,\partial_i \ln m_{\boldsymbol{\theta}}(x)| \le C(1+|x|^l)$, then $|\,\partial_i \ln m^*_{\boldsymbol{\theta}}(x)| \le C^*(1+|x|^l)$.

(e4). Let $E_m[\cdot]$ denotes the expectation taken with respect to $m_\theta(x)$. Then for some positive integer l,

$$E_m\big[|X|^l\big] < \infty \qquad \text{implies} \qquad E^*\big[|X|^l\big] < \infty$$

Remark All the derivatives shown in the above proposition can be replaced by higher order derivatives up to order three. The first order derivative ∂_i is used for notation simplicity.

Proof: By Proposition B.2.5, it is valid to interchange the integration and differentiation when finding (higher order) derivatives for $m_{\boldsymbol{\theta}}^*(x)$. We will use this result in this proof whenever needed without explicitly mentioning it.

For $(e1)$, by the change of variables $x - t = u$ and $t = v$,

$$\begin{aligned}\left| \partial_i \int_{\mathbb{R}} m_{\boldsymbol{\theta}}^*(x)\,\mathrm{d}x \right| &= \left| \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\, e^{-(x-t)^2/2}\ \partial_i m_{\boldsymbol{\theta}}(t)\,\mathrm{d}t\ \mathrm{d}x \right| \\ &\leq \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}}\, e^{-u^2/2} \left| \partial_i m_{\boldsymbol{\theta}}(v) \right| \mathrm{d}v\,\mathrm{d}u \ < \infty\end{aligned}$$

For $(e2)$, assume $\left| \partial_i m_{\boldsymbol{\theta}}(t) \right| \leq K_i(\boldsymbol{\theta})$. Then

$$\begin{aligned}\left| \partial_i\, m_{\boldsymbol{\theta}}^*(x) \right| &\leq \int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} \left| \partial_i\, m_{\boldsymbol{\theta}}(t) \right| \mathrm{d}t \\ &= \int_{\mathbb{R}} \frac{e^{-z^2/2}}{\sqrt{2\pi}} \left| \partial_i\, m_{\boldsymbol{\theta}}(z + x) \right| \mathrm{d}z \\ &\leq \int_{\mathbb{R}} \frac{e^{-z^2/2}}{\sqrt{2\pi}}\, K_i(\boldsymbol{\theta})\,\mathrm{d}z \\ &= K_i(\boldsymbol{\theta})\end{aligned}$$

For $(e3)$, notice that $|\,\partial_i \ln m_{\boldsymbol{\theta}}(x)| \leq C(1 + |x|^l)$ is equivalent to $|\,\partial_i m_{\boldsymbol{\theta}}(x)| \leq C(1 + |x|^l) m_{\boldsymbol{\theta}}(x)$, therefore

$$\begin{aligned}\partial_i\, m_{\boldsymbol{\theta}}^*(x) &= \int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}\, \partial_i m_{\boldsymbol{\theta}}(t)\,\mathrm{d}t \\ &\leq \int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}\, C(1 + |t|^l)\, m_{\boldsymbol{\theta}}(t)\,\mathrm{d}t \\ &= C\, m_{\boldsymbol{\theta}}^*(t) \ + \ C \ \int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}}\, |t|^l\, m_{\boldsymbol{\theta}}(t)\,\mathrm{d}t \end{aligned} \tag{B.22}$$

Consider the case where $x \gg 1$. From the result to be shown (Proposition B.2.10), $E[|X|^p]$ is a finite continuous function of $\boldsymbol{\theta}$ for any positive integer l. Let $M = \sup_{\boldsymbol{\theta} \in B_a} E[|X|^{10}]$ and Δ to be an arbitrary small positive constant, i.e., $0 < \Delta \ll 1$. It is easy to see, when $x \gg 1$, $e^{-\Delta x}\, x^l < 1$ for any l.

$$
\begin{aligned}
\int_{t \geq Mx} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} |t|^l\, m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t &= \int_{t \geq Mx} \frac{e^{-(x-t)^2/2 + \Delta t}}{\sqrt{2\pi}} \left(e^{-\Delta t}\, |t|^l\right) m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&< \int_{t \geq Mx} \frac{e^{-((x+\Delta)-t)^2/2 + \Delta x + \Delta^2/2}}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&\leq e^{-(M-1)^2 x^2/2 + \Delta M x} \cdot \int_{t \geq Mx} \frac{1}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t
\end{aligned}
$$

On the other hand,

$$
\begin{aligned}
\int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t &> \int_{t \leq 3x} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&> e^{-4x^2/2} \cdot \int_{t \leq 3x} \frac{1}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&> e^{-(M-1)^2 x^2/2 + \Delta M x} \cdot \int_{t \geq Mx} \frac{1}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t
\end{aligned}
$$

for large x and M. Therefore, the second term in (B.22) satisfies

$$
\begin{aligned}
\int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} |t|^l\, m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t &= \int_{t \leq Mx} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} |t|^l\, m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t + \int_{t \geq Mx} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} |t|^l\, m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&< M^l x^l \int_{t \leq Mx} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t + \int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&< \left(M^l x^l + 1\right) \cdot \int_{\mathbb{R}} \frac{e^{-(x-t)^2/2}}{\sqrt{2\pi}} m_{\boldsymbol{\theta}}(t)\, \mathrm{d}t \\
&= \left(M^l x^l + 1\right) m^*_{\boldsymbol{\theta}}(x)
\end{aligned}
$$

Plug this upper bound in (B.22) and $(e3)$ follows. The case where $x \ll -1$ can be proved similarly.

For $(e4)$,

$$E^*[|X|^l] = \int_{\mathbb{R}} |x|^l\, m^*_{\boldsymbol{\theta}}(x)\,\mathrm{d}x = \int_{\mathbb{R}} |x|^l \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2} m_{\boldsymbol{\theta}}(t)\,\mathrm{d}t\ \mathrm{d}x$$

Use the following substitution,

$$x - t = u \quad \text{and} \quad t = v,$$

$$\begin{aligned}
E^*[|X|^l] &= \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-u^2/2}\, |v-u|^l\, m_{\boldsymbol{\theta}}(v)\,\mathrm{d}v\,\mathrm{d}u \\
&\le \int_{\mathbb{R}}\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-(x-t)^2/2}\, l!(|v|^l + |u|^l)\, m_{\boldsymbol{\theta}}(v)\,\mathrm{d}v\,\mathrm{d}u \\
&= \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} l!\,|u|^l\, e^{-u^2/2}\,\mathrm{d}u \ +\ l! \int_{\mathbb{R}} |v|^l\, m_{\boldsymbol{\theta}}(v)\,\mathrm{d}v\,\mathrm{d}u \\
&< \infty
\end{aligned}$$

■

Proposition B.2.6 implies, to study the derivation of the Taylor expansion, we need only to focus on $m_{\boldsymbol{\theta}}(x)$. As the absolute continuity of $m_{\boldsymbol{\theta}}(x)$ with respect to $\boldsymbol{\theta}$ has been shown in the proof of Proposition B.2.5, it is left to prove

$$\frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}} m_{\boldsymbol{\theta}}(x) \le C(1 + |x|^l)\, m_{\boldsymbol{\theta}}(x) \tag{B.23}$$

for large $|x|$. Recall that the derivatives to be considered are those with respect to the parameter $\boldsymbol{\theta} = (\mu, \beta, \rho, \nu, \alpha)$ and are of order up to 3. Before proving (B.23), let us first present some preliminary results.

Lemma B.2.7: For sufficiently large $|x|$ and fixed (y, w),

$$\int_{\mathbb{R}} s^q g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}s \le \Psi_{3,q}(h, \boldsymbol{\theta})\, (1 + |w|^q) \int_{\mathbb{R}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}s \tag{B.24}$$

and

$$\int_{\mathbb{R}} s^q g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}s \;\le\; M^q\, |x|^q \int_{\mathbb{R}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}s \tag{B.25}$$

where q is some positive integer and $\Psi_{3,q}(h, \boldsymbol{\theta})$ is a continuous function of $\boldsymbol{\theta}$.

Proof Consider the integral $\iiint_{\mathbb{D}} s^q g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s$ and the reparametrization when $x \gg 1$:

$$\mu = x\epsilon, \quad y = x t_1, \quad w = x t_2 \quad \text{and} \quad s = x t_3.$$

Let $\mathbb{D}_s$ denote the region $\{0 \le t_1 \le (1 - e^{-h}) t_2,\ 0 \le t_2 \text{ and } 0 \le t_3\}$. One has

$$\begin{aligned}
&\iiint_{\mathbb{D}} s^q\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= \iiint_{\mathbb{D}_s} x^{q+3}\, t_3^q\, g(x, t_1, t_2, t_3; \boldsymbol{\theta})\, \mathrm{d}t_1\, \mathrm{d}t_2\, \mathrm{d}t_3 \\
&= x^{q+\nu+3/2} \iiint_{\mathbb{D}_s} t_3^q\, \frac{f_{Y,W}(t_1, t_2)}{\sqrt{2\pi}\, \sqrt{(1 - e^{-h}) t_3 + t_1}}\, \frac{\alpha^\nu}{\Gamma(\nu)}\, t_3^{\nu-1}\, e^{-\alpha\, x\, t_3} \\
&\qquad \exp\left\{ -\, x\, \frac{\big(1 - \epsilon - \beta((1 - e^{-h}) t_3 + t_1) - \rho t_2\big)^2}{2((1 - e^{-h}) t_3 + t_1)} \right\} \mathrm{d}t_1\, \mathrm{d}t_2\, \mathrm{d}t_3.
\end{aligned}$$

Isolate the integration with respect to t_3,

$$\iiint_{\mathbb{D}} s^q \, g(x, y, w, s; \boldsymbol{\theta}) \, \mathrm{d}y \ \mathrm{d}w \ \mathrm{d}s$$

$$= x^{q+\nu+3/2} \int_0^{\infty} \int_0^{(1-e^{-h})t_2} \frac{\alpha^{\nu}}{\sqrt{2\pi}\,\Gamma(\nu)} f_{Y,W}(t_1, t_2)$$

$$\int_0^{\infty} \frac{t_3^{q+\nu-1}}{\sqrt{(1-e^{-h})t_3 + t_1}} \exp\Big\{ - x \, h(t_1, t_2, t_3) \Big\} \, \mathrm{d}t_3 \ \mathrm{d}t_1 \, \mathrm{d}t_2$$

where

$$h(t_1, t_2, t_3) = \frac{\big(1 - \epsilon - \beta((1 - e^{-h})t_3 + t_1) - \rho t_2\big)^2}{2((1 - e^{-h})t_3 + t_1)} + \alpha t_3.$$

According to the Laplace method (cf [27], Section 2.4), for large x and fixed (t_1, t_2), the main contribution of the integral

$$I_1(x) = \int_0^{\infty} \frac{t_3^{q+\nu-1}}{\sqrt{(1-e^{-h})t_3 + t_1}} \exp\Big\{ - x \, h(t_1, t_2, t_3) \Big\} \, \mathrm{d}t_3.$$

comes from the integration in the neighborhood of the locally minimizing values of $h(t_1, t_2, t_3)$ (if any) over t_3. To find the critical numbers, one solves $\partial h / \partial t_3 = 0$. It turns out that this is a quadratic equation with one root negative and the other root given by

$$\begin{aligned} t_3^* &= \frac{1}{(e^h - 1)^2 (e^h(2\alpha + \beta^2) - \beta^2)} \cdot \Big(- e^h (e^h - 1)(e^h(2\alpha + \beta^2) - \beta^2) \, t_1 + \\ &\qquad \sqrt{e^{2h}(e^h - 1)^3 (e^h(2\alpha + \beta^2) - \beta^2)(\epsilon - 1 + \rho t_2)^2} \Big) \\ &= -\frac{e^h}{e^h - 1} t_1 + \frac{e^h}{\sqrt{(e^h - 1)(e^h(2\alpha + \beta^2) - \beta^2)}} \big| -1 + \epsilon + \rho \, t_2 \big| \end{aligned} \tag{B.26}$$

Evaluating $h''(t_1, t_2, t_3)$ at t_3^*, one gets

$$h''(t_1, t_2, t_3^*) = \frac{e^{-2h}\sqrt{(e^h - 1)(e^h(2\alpha + \beta^2) - \beta^2)^3}}{|-1 + \epsilon + \rho\, t_2|} > 0$$

We know t_3^* is in fact a global minimum of $h(t_1, t_2, t_3)$ which in turn maximizes $e^{-xh(t_1,t_2,t_3)}$. If $t_3^* > 0$, by the Laplace method,

$$I_1(x) \approx e^{-xh(t_1,t_2,t_3^*)} \frac{(t_3^*)^{q+\nu-1}}{\sqrt{(1-e^{-h})t_3^* + t_1}} \sqrt{\frac{2\pi}{xh''(t_1, t_2, t_3^*)}}$$

as $x \to \infty$. It is clear that we can isolate the integral over s for

$$\iiint_{\mathbb{D}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s$$

and perform the same reparameterization to get

$$\begin{aligned}
&\iiint_{\mathbb{D}} g((x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= x^{\nu+3/2} \int_0^\infty \int_0^{(1-e^{-h})t_2} \frac{\alpha^\nu}{\sqrt{2\pi}\,\Gamma(\nu)}\, f_{Y,W}(t_1, t_2)\, I_0(x)\ \mathrm{d}t_1\, \mathrm{d}t_2
\end{aligned}$$

where

$$I_0(x) = \int_0^\infty \frac{t_3^{\nu-1}}{\sqrt{(1-e^{-h})t_3 + t_1}} \exp\Big\{ -x\, h(t_1, t_2, t_3) \Big\}\, \mathrm{d}t_3.$$

Applying the Laplace method:

$$I_0(x) \approx e^{-xh(t_1,t_2,t_3^*)} \frac{(t_3^*)^{\nu-1}}{\sqrt{(1-e^{-h})t_3^* + t_1}} \sqrt{\frac{2\pi}{xh''(t_1, t_2, t_3^*)}}.$$

Therefore,

$$
\begin{aligned}
I_1(x) &\approx I_0(x)\,(t_3^*)^q \\
&= I_0(x)\Big(-\frac{e^h}{e^h-1}\,t_1 \;+\; \frac{e^h\,\big|-1+\epsilon+\rho\,t_2\big|}{\sqrt{(e^h-1)(e^h(2\alpha+\beta^2)-\beta^2)}}\Big)^q \\
&\leq q!\,I_0(x)\Big[\Big(\frac{e^h}{e^h-1}\,t_1\Big)^q \;+\; \Big(\frac{e^h\,\big|-1+\epsilon+\rho\,t_2\big|}{\sqrt{(e^h-1)(e^h(2\alpha+\beta^2)-\beta^2)}}\Big)^q\Big] \\
&\leq q!\,I_0(x)\Big[t_2^q \;+\; \Big(\frac{e^h\,\big|-1+\epsilon+\rho\,t_2\big|}{\sqrt{(e^h-1)(e^h(2\alpha+\beta^2)-\beta^2)}}\Big)^q\Big] \\
&\leq I_0(x)\,\Psi_{3,q}(h,\boldsymbol{\theta})\,(1+t_2^q)
\end{aligned}
$$

The second inequality holds because $t_1 \leq (1-e^{-h})t_2$. If we reparameterize (t_1,t_2,t_3) back to (y,w,s), we see that (B.24) holds.

If $t_3^* < 0$, then there is no minimizer of $h(t_1,t_2,t_3)$ since t_3 is defined on $[0,\infty)$. This implies for the corresponding (y,w) value, $g(x,y,w,s;\boldsymbol{\theta})$ decreases exponentially fast over s when $s \gg 1$. Using the similar tail estimate approach in the proof of Proposition B.2.6, one can show

$$
\int_{\mathbb{R}} g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}s \;>\; \int_{s>Mx} s^q\,g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}s
$$

for $x \gg 1$ and large M, so that

$$
\begin{aligned}
\int_{\mathbb{R}} s^q g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}s \;&=\; \int_{s\leq Mx} s^q g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}s \;+\; \int_{s>Mx} s^q g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}s \\
&< (M^q x^q + 1)\int_{\mathbb{R}} g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}s\,.
\end{aligned}
$$

Therefore the bound (B.25) holds.

When $x \ll -1$ and $x \to -\infty$, one can consider the following reparameterization

$$\mu = -x\epsilon, \quad y = -xt_1, \quad w = -xt_2 \quad \text{and} \quad s = -xt_3.$$

with the corresponding

$$h(t_1, t_2, t_3) = \frac{\left(1 + \epsilon + \beta((1 - e^{-h})t_3 + t_1) + \rho t_2\right)^2}{2((1 - e^{-h})t_3 + t_1)} - \alpha t_3.$$

By the same arguments above, one can show the bounds (B.24) and (B.25) are valid.

■

Next we consider the finiteness of the exponential moments of Y, W, S and X. This result will come handy later when estimating the tail mass of some integrals. We first state a known result related to moments of functions of Lévy process.

Lemma B.2.8: **(Sato 1999, Theorem 25.3)** Let Z_t be a Lévy process with Lévy measure $w(x)$. If $h(x)$ is a submultiplicative, locally bounded, measurable function on $\mathbb{R}$, then $E[h(Z_1)]$ is finite if and only if

$$\int_{|x|>1} h(x) w(x) \, \mathrm{d}x < \infty.$$

Proposition B.2.9: For random variables Y, W and S defined in (B.11), one has

(m1). For $\forall c_1 < \alpha$ $E[e^{c_1 W}] < \infty$ and $E[e^{\frac{c_1}{1-e^{-h}} Y}] < \infty$.

(m2). $E[e^{c_2(Y+W)}] < \infty$ if $c_2 < \frac{\alpha}{2-e^{-h}}$.

(m3). $E[e^{c_3 S}] < \infty$ for $c_3 < \alpha$.

Proof: Since W is a Compound Poisson random variable with Lévy measure $w(x) = \nu\alpha e^{-\alpha x}$, by Lemma B.2.8, $E[e^{c_1 W}] < \infty$ for $\forall c_1 < \alpha$. From the definition (B.11), $Y \leq (1 - e^{-h})W$ a.s., thus $E[e^{\frac{c_1}{1-e^{-h}}Y}] \leq E[e^{c_1 W}] < \infty$. Further, if c_2 is chosen in such a way that $c_2(2 - e^{-h}) < \alpha$, then $E[e^{c_2(Y+W)}] < \infty$ holds. Finally, $S \sim \text{Gamma}(\nu, \alpha)$ implies $E[e^{c_3 S}] < \infty$ for $\forall c_3 < \alpha$.

■

Proposition B.2.10: There exists a positive number b such that $E_m[e^{b|X|}] < \infty$. As a consequence, all polynomial moments of X are finite.

Proof: Recall if $\zeta \sim \mathrm{N}(\tilde{\mu}, \tilde{\sigma}^2)$, then

$$E[e^{b\,\zeta}] = \tilde{\sigma}\ \exp\big(b\tilde{\mu} + \frac{b^2}{2}\tilde{\sigma}^2\big) \quad \text{and} \quad E[e^{-b\,\zeta}] = \tilde{\sigma}\ \exp\big(-b\tilde{\mu} + \frac{b^2}{2}\tilde{\sigma}^2\big)$$

Since

$$E_m[e^{b|X|}] < E_m[e^{bX}] + E_m[e^{-bX}].$$

for $b > 0$, consider those two terms separately,

$$\begin{aligned}
E_m[e^{bX}] &= E\Big[E[e^{bX}|Y, W, S]\Big] \\
&= E\Big[\exp\big(b\mu + b\beta((1-e^{-h})S + Y) + b\rho W\big) \\
&\qquad \cdot \exp\big(\frac{b^2}{2}((1-e^{-h})S + Y)\big) \cdot ((1-e^{-h})S + Y)\Big].
\end{aligned}$$

It is clear that we need only to consider the finiteness of $E_m[e^{bX}]$ when $\beta > 0$.

Grouping those terms which increase as (Y, W, S) increases, one finds to guarantee $E_m[e^{bX}] < \infty$, it is sufficient, by independence of S,Y, to show

$$E\big[\exp((1-e^{-h})(\tfrac{b^2}{2} + b\beta)\, S)\big] < \infty$$

and

$$E\big[\exp((\tfrac{b^2}{2} + b\beta)\, Y)\big] \le E\big[\exp((1-e^{-h})(\tfrac{b^2}{2} + b\beta)\, W)\big] < \infty.$$

The term $b\rho W$ is dropped because $\rho < 0$. By Proposition (B.2.9), the expectations above are finite if the following inequality holds,

$$(1-e^{-h})(\frac{b^2}{2} + b\beta) < \alpha. \tag{B.27}$$

Solving this quadratic inequality with respect to b, we find the roots are given by

$$b^* = -\beta \pm \sqrt{\beta^2 + \frac{2\alpha}{1-e^{-h}}}.$$

Since one of the roots is positive, then

$$0 < b \le -\beta + \sqrt{\beta^2 + \frac{2\alpha}{1-e^{-h}}}.$$

gives the solution to (B.27). One finds there must exist some positive b where $E_m[e^{bX}] < \infty$.

Now we turn to examine $E_m[e^{-bX}]$. By the similar arguments above, to guarantee $E_m[e^{-bX}] < \infty$, consider

$$
\begin{aligned}
E_m[e^{-bX}] &= E\Big[E[e^{-bX}|Y, W, S]\Big] \\
&= E\Big[\exp\big(-b\mu - b\beta((1-e^{-h})S + Y) - b\rho W\big) \\
&\qquad \cdot \exp\big(\frac{b^2}{2}((1-e^{-h})S + Y)\big) \cdot ((1-e^{-h})S + Y)\Big].
\end{aligned}
$$

To make the expectation finite, one ends up solving the following two inequalities:

- $(1-e^{-h}) \cdot \big(\frac{b^2}{2} - b\beta\big) < \alpha$
- $(1-e^{-h}) \cdot \big(\frac{b^2}{2} - b\beta\big) - b\rho < \alpha$

for $\beta < 0$. It is not difficult to see both inequalities contain positive solutions, therefore $E_m[e^{-bX}] < \infty$ for some $b > 0$. And we can conclude that

$$
E_m[e^{b|X|}] < E_m[e^{bX}] + E_m[e^{-bX}] < \infty \qquad \text{for some} \quad b > 0.
$$

$\blacksquare$

Remark Since S is a Gamma r.v. and W is a compound Poisson r.v. with jump sizes following the exponential distribution, the exponential moments of S and W, if exist, are continuous functions of the parameters. Therefore, the polynomial moments of $|X|$ are bounded by finite continuous functions of $\boldsymbol{\theta}$.

The last result gives an upper bound for the integral

$$
\iiint_{\mathbb{D}} w^r g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y \ \mathrm{d}w \ \mathrm{d}s
$$

Lemma B.2.11: The following inequality holds for sufficiently large $|x|$

$$\iiint_{\mathbb{D}} w^r\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \ < \qquad \text{(B.28)}$$
$$\Psi_{4,r}(h,\boldsymbol{\theta})\, |x|^r \iiint_{\mathbb{D}} g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s$$

Proof: Recall the definition of $g(x,y,w,s;\boldsymbol{\theta})$ from (B.18):

$$g(x,y,w,s;\boldsymbol{\theta}) \ = \ \frac{f_{Y,W,S}(y,w,s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s+y}} \exp\left\{-\frac{\big(x-\mu-\beta((1-e^{-h})s+y)-\rho w\big)^2}{2((1-e^{-h})s+y)}\right\}$$

When $x \gg 1$, for a given large positive constant $M > 0$, consider the decomposition of the integral on the left-hand side of (B.28),

$$\begin{aligned}
&\iiint_{\mathbb{D}} w^r\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= \Big(\int_0^{\infty}\int_0^{\frac{M}{|\rho|}x}\int_0^{(1-e^{-h})w} + \int_0^{\infty}\int_{\frac{M}{|\rho|}x}^{\infty}\int_0^{(1-e^{-h})w}\Big)\, w^r\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\le \int_0^{\infty}\int_0^{\frac{M}{|\rho|}x}\int_0^{(1-e^{-h})w} \frac{M^r}{|\rho|^r}\, x^r\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\quad + \int_0^{\infty}\int_{\frac{M}{|\rho|}x}^{\infty}\int_0^{(1-e^{-h})w} w^r\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\le \frac{M^r}{|\rho|^r}\, x^r \iiint_{\mathbb{D}} g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \ + \iiint_{\mathbb{D}_x} w^r\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s
\end{aligned}$$

where

$$\mathbb{D}_x \triangleq \left\{0 \le y \le (1-e^{-h})w,\ \frac{M}{|\rho|}x < w \ \text{ and } \ 0 \le s\right\} \tag{B.29}$$

To prove (B.28), it suffices to show

$$\iiint_{\mathbb{D}_x} w^r\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \ < \ \iiint_{\mathbb{D}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \tag{B.30}$$

for sufficiently large x. Since $S \sim \text{Gamma}(\nu, \alpha)$ with $\nu > \frac{7}{2}$,

$$\iiint_{\mathbb{D}_x} \frac{f_{Y,W,S}}{\sqrt{(1-e^{-h})s + y}} < \iiint_{\mathbb{D}_x} \frac{f_{Y,W,S}}{\sqrt{(1-e^{-h})s}} < \infty$$

Therefore, by Cauchy-Schwartz inequality,

$$\begin{aligned}
&\iiint_{\mathbb{D}_x} w^r\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
\le &\iiint_{\mathbb{D}_x} \sqrt{\frac{f_{Y,W,S}(y, w, s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s + y}}} \cdot \\
&\quad w^r \sqrt{\frac{f_{Y,W,S}(y, w, s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s + y}}}\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
\le &\left(\iiint_{\mathbb{D}_x} \frac{f_{Y,W,S}(y, w, s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s + y}}\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \right)^{1/2} \cdot \\
&\quad \left(\iiint_{\mathbb{D}_x} w^{2r} \frac{f_{Y,W,S}(y, w, s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s + y}}\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \right)^{1/2} \\
= &\, R_1^{1/2} \cdot R_2^{1/2}
\end{aligned}$$

To get the upper bound of $\iiint_{\mathbb{D}_x} w^r\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s$, let us first consider R_1. Choose $c_1 < \alpha$, it is not difficult to see

$$
\begin{aligned}
R_1 &\leq \iiint_{\mathbb{D}_x} \frac{f_{Y,W,S}(y,w,s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s}}\,\mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= \frac{1}{\sqrt{2\pi(1-e^{-h})}} \iiint_{\mathbb{D}_x} \frac{1}{\sqrt{s}}\, f_{Y,W,S}(y,w,s)\ \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= \frac{E[S^{-1/2}]}{\sqrt{2\pi(1-e^{-h})}}\,\mathbb{P}\big(W > \frac{M}{|\rho|}x\big) \\
&\leq \frac{E[S^{-1/2}]\,E[e^{c_1 W}]}{\sqrt{2\pi(1-e^{-h})}}\ \exp\big(-c_1\frac{M}{|\rho|}x\big)
\end{aligned}
$$

The last line holds since the upper bound of $\mathbb{P}\big(W > \frac{M}{|\rho|}x\big)$ can be derived from the result $(m3)$ of Proposition B.2.9 and the Markov inequality.Next we consider R_2. Choose the same c_1 as in R_1 and another constant Δ such that $0 < \Delta \ll \frac{\alpha}{2}$,

$$
\begin{aligned}
R_2 &= \iiint_{\mathbb{D}_x} e^{-\Delta w}\, w^{2r}\, \frac{e^{\Delta w} f_{Y,W,S}(y,w,s)}{\sqrt{2\pi}\,\sqrt{(1-e^{-h})s+y}}\,\mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\leq \frac{1}{\sqrt{2\pi(1-e^{-h})}} \iiint_{\mathbb{D}_x} \frac{1}{\sqrt{s}}\, e^{\Delta w}\ f_{Y,W,S}(y,w,s)\,\mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&= \frac{E\big[S^{-1/2}\big]}{\sqrt{2\pi(1-e^{-h})}} \cdot E\big[e^{\Delta W}\mathbb{I}_{\{W \geq \frac{M}{|\rho|}x\}}\big] \\
&\leq \frac{E\big[S^{-1/2}\big]}{\sqrt{2\pi(1-e^{-h})}} \cdot \sqrt{E\big[e^{2\Delta W}\big]}\,\sqrt{E\big[\mathbb{I}_{\{W \geq \frac{M}{|\rho|}x\}}\big]} \\
&< \frac{E\big[S^{-1/2}\big]}{\sqrt{2\pi(1-e^{-h})}} \cdot \sqrt{E[e^{2\Delta W}]\,E[e^{c_1 W}]}\ \exp\big(-\frac{c_1 M}{2|\rho|}x\big)
\end{aligned}
$$

The second line holds because $e^{-\Delta w} w^{2r} < 1$ for large w. Therefore,

$$
R_2^{1/2} \leq \left(\frac{E\big[S^{-1/2}\big]\,\sqrt{E[e^{2\Delta W}]\,E[e^{c_1 W}]}}{\sqrt{2\pi(1-e^{-h})}}\right)^{1/2}\ \exp\big(-\frac{c_1 M}{4|\rho|}x\big) \tag{B.31}
$$

Combining the bounds on R_1 and R_2,

$$\iiint_{\mathbb{D}_x} w^r\, g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \tag{B.32}$$
$$< \frac{E\big[S^{-1/2}\big]\,E[e^{2\Delta W}]^{1/4}}{\sqrt{2\pi(1-e^{-h})}}\big(E[e^{c_1 W}]\big)^{3/4}\cdot\exp\big(-c_1\frac{3M}{4|\rho|}x\big)$$

Next we consider finding the lower bound of the $\iiint_{\mathbb{D}} g(x,y,w,s;\boldsymbol{\theta})\,\mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s$. Define a subset $\tilde{\mathbb{D}}_x$ of $\mathbb{D}$ by

$$\tilde{\mathbb{D}}_x : \Big\{0 < w \le U_w \quad \text{where } U_w \text{ is the median of the distribution of} \quad W$$
$$\text{and} \quad s > \frac{\delta}{1-e^{-h}}\,x \quad \text{for} \quad \delta > 0\Big\}$$

Since $Y \le (1-e^{-h})W$ a.s. by there definitions, $\tilde{\mathbb{D}}_x$ is bounded on (y,w) and such that

$$\mathbb{P}\big((Y,W)\in\tilde{\mathbb{D}}_x\big) = \tfrac{1}{2}.$$

Notice that $x \gg U_w$, one has

$$\frac{\Big(x-\mu-\beta\big((1-e^{-h})s+y\big)-\rho w\Big)^2}{2((1-e^{-h})s+y)} \le \frac{\big(x-\beta(1-e^{-h})s\big)^2}{(1-e^{-h})s} + \frac{\big(\mu-\beta y-\rho w\big)^2}{(1-e^{-h})s}.$$

Maximize the first term on the righ-hand side with respect to s gives

$$3\beta(1-e^{-h})s = x$$

Then it is clear that when $(y, w, s) \in \tilde{\mathbb{D}}_x$

$$\frac{\big(x-\beta(1-e^{-h})s\big)^2}{(1-e^{-h})s} \quad \leq \quad \begin{cases} \dfrac{(1-\beta\delta)^2}{\delta}\, x \\ \\ \dfrac{4\beta}{3}\, x \end{cases}$$

Therefore

$$\frac{\Big(x-\mu-\beta\big((1-e^{-h})s+y\big)-\rho w\Big)^2}{2((1-e^{-h})s+y)} \leq \tilde{C}x + \varepsilon_x$$

where $\tilde{C} = \min\big((1-\beta\delta)^2/\delta,\, 4\beta/3\big)$ and $\varepsilon_x \to 0$ as $x \to \infty$. One then finds

$$\begin{aligned}
&\iiint_{\mathbb{D}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\quad\geq \iiint_{\tilde{\mathbb{D}}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\quad\geq \frac{1}{\sqrt{2\pi}} \iiint_{\tilde{\mathbb{D}}} \frac{f_{Y,W,S}(y, w, s)}{\sqrt{(1-e^{-h})s+y}} \exp\Big(-\tilde{C}x - \varepsilon_x\Big)\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\quad\geq \frac{\exp\big(-\tilde{C}x - \varepsilon_x\big)}{\sqrt{2\pi}} \iiint_{\tilde{\mathbb{D}}} \frac{f_{Y,W,S}(y, w, s)}{\sqrt{2s}}\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\quad\geq \frac{\exp\big(-\tilde{C}x - \varepsilon_x\big)}{2\pi} \left[\frac{1}{2}\int_{\frac{\delta}{1-e^{-h}}x}^{\infty} \frac{\alpha^\nu}{\Gamma(\nu)}\, s^{\nu-3/2} e^{-\alpha s}\, \mathrm{d}s\right] \\
&\quad\geq \frac{\exp\big(-\tilde{C}x - \varepsilon_x\big)}{4\pi} \exp\Big(-\frac{\alpha\delta}{1-e^{-h}}x\Big)
\end{aligned} \tag{B.33}$$

Compared two inequalities given by (B.33) and (B.32), it is always possible to find such an $\boldsymbol{M}$, which depends on the parameters, c_1, δ and h, that (B.30) holds. The case where $x \ll -1$ can be shown similarly, therefore the conclusion (B.28) is justified. ■

At last, we give the tail estimate for $\partial_i \ln m_{\boldsymbol{\theta}}(x)$.

Lemma B.2.12: For the model density $m_{\boldsymbol{\theta}}(x)$ in the Γ-OU BN-S model. If $\nu > 7/2$, then the following bound holds with $p \leq 3$

$$\frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}} m_{\boldsymbol{\theta}}(x) \leq C(h, \boldsymbol{\theta})\,(1 + |x|^l) m_{\boldsymbol{\theta}}(x) \tag{B.34}$$

for some constant $C(h, \boldsymbol{\theta})$ and integer l which depend on $(p_1, \ldots, p_5)$.

Proof: From Proposition B.2.2,

$$\frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}} m_{\boldsymbol{\theta}}(x) \;<\; \iiint_{\mathbb{D}} \sum_{l,q,r} \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |x|^l\, s^q\, w^r\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s\,.$$

By Lemma B.2.7 and Lemma B.2.11,

$$\begin{aligned} &\iiint_{\mathbb{D}} \sum_{l,q,r} \Psi_{2,\mathrm{i}}(h, \boldsymbol{\theta})\, |x|^l\, s^q\, w^r\, g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\ &\qquad < \sum_{l,q,r} \Psi_{\mathrm{i}}(h, \boldsymbol{\theta}) |x|^{l+q+r} \iiint_{\mathbb{D}} g(x, y, w, s; \boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \end{aligned}$$

for sufficiently large $|x|$. And we know the boundedness of $\frac{\partial^p}{\partial_1^{p_1} \cdots \partial_5^{p_5}} m_{\boldsymbol{\theta}}(x)$ in the proof of the result (B.34) is justified. ■

The previous discussions have provided the required details to justify the interchange of differentiation and integration for deriving the Taylor expansion. For illustration purpose, let us apply the steps discussed above to verify:

$$\frac{\partial}{\partial\beta\,\partial\alpha}\int_{\mathbb{R}}\big(e^{-\delta^n}-2\big)\,m^*_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \;=\int_{\mathbb{R}}\frac{\partial}{\partial\beta\,\partial\alpha}\big(e^{-\delta^n}-2\big)\,m^*_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \tag{B.35}$$

Remark As shown in Proposition B.2.6, it is sufficient to focus on $m_{\boldsymbol{\theta}}(x)$ instead of the kernel convolved density $m^*_{\boldsymbol{\theta}}(x)$ in order to apply Lemma B.2.3, we thus replace the $m^*_{\boldsymbol{\theta}}(x)$ by $m_{\boldsymbol{\theta}}(x)$ in the expression (B.35) for notation simplicity.

Recall the expression (3.11) in Lemma 3.2.3, define $\rho_{25}(\boldsymbol{\theta})$ by

$$\begin{aligned}
\rho_{25}(\boldsymbol{\theta}) &\triangleq \frac{\partial}{\partial\beta\,\partial\alpha}\big(e^{-\frac{f^*}{m_{\boldsymbol{\theta}}}+1}-2\big)\,m_{\boldsymbol{\theta}}(x)\\
&= e^{-\frac{f^*}{m_{\boldsymbol{\theta}}}+1}\cdot\frac{f^*(x)^2}{m_{\boldsymbol{\theta}}(x)^2}\cdot\,\partial_2\ln m_{\boldsymbol{\theta}}(x)\cdot\,\partial_5\ln m_{\boldsymbol{\theta}}(x)\cdot m_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \qquad\qquad \text{(B.36)}\\
&\qquad + e^{-\frac{f^*}{m_{\boldsymbol{\theta}}}+1}\cdot\frac{f^*(x)}{m_{\boldsymbol{\theta}}(x)}\cdot\,\partial_{25}m_{\boldsymbol{\theta}}(x)\,\mathrm{d}x \;+\;(e^{-\frac{f^*}{m_{\boldsymbol{\theta}}}+1}-2)\,\partial_{25}m_{\boldsymbol{\theta}}(x)\,\mathrm{d}x
\end{aligned}$$

(1) First we show $\partial_2 m_{\boldsymbol{\theta}}(x)$, $\partial_5 m_{\boldsymbol{\theta}}(x)$ and $\partial_{25} m_{\boldsymbol{\theta}}(x)$ can be computed for $\beta\in B_{a,2}\triangleq[\beta^s-a,\beta^s+a]$ and $\alpha\in B_{a,5}=[\alpha^s-a,\alpha^s+a]$ by interchanging differentiation and integration. From Proposition B.2.1 and B.2.2,

$$\begin{aligned}
\partial_2\,g(x,y,w,s;\boldsymbol{\theta}) &= g(x,y,w,s;\boldsymbol{\theta})\,\Big(x-\mu-\beta((1-e^{-h})s+y)-\rho w\Big)\\
\partial_5\,g(x,s,y,w;\boldsymbol{\theta}) &= g(x,s,y,w;\boldsymbol{\theta})\,\Big(-\frac{\nu}{\alpha}s-w+\frac{\nu}{h}\Big)
\end{aligned}$$

with their absolute values bounded by

$$
\begin{aligned}
|\,\partial_2\, g(x,y,w,s;\boldsymbol{\theta})| &\le g(x,y,w,s;\boldsymbol{\theta})\, \Big(|x| + |\mu| + |\beta|((1-e^{-h})s + y) + |\rho| w\Big) \\
|\,\partial_5\, g(x,s,y,w;\boldsymbol{\theta})| &\le g(x,s,y,w;\boldsymbol{\theta})\, \left(\frac{\nu}{\alpha} s + w + \frac{\nu}{h}\right)
\end{aligned}
$$

It is easy to see $g(x,y,w,s;\boldsymbol{\theta})$ is absolutely continuous with respect to β and α since the derivatives exist and are bounded on $\boldsymbol{B}_{a,2}$ and $\boldsymbol{B}_{a,5}$ for all finite $|x|$. Notice that we can bound $g(x,y,w,s;\boldsymbol{\theta})$ by

$$
\begin{aligned}
g(x,y,w,s;\boldsymbol{\theta}) &\le \frac{1}{\sqrt{(1-e^{-h})s}}\, f_{Y,W,S}(y,w,s) \qquad \text{or} \\
g(x,y,w,s;\boldsymbol{\theta}) &\le \frac{1}{\sqrt{y}}\, f_{Y,W,S}(y,w,s)
\end{aligned}
$$

Then it is easy to check, for fixed x:

$$
\begin{aligned}
&\Big|\,\partial_2 m_{\boldsymbol{\theta}}(x)\Big| \;\le\; \iiint_{\mathbb{D}} \Big|\,\partial_2\, g(x,y,w,s;\boldsymbol{\theta})\Big|\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\le \iiint_{\mathbb{D}} \Big(|x| + |\mu| + |\beta|((1-e^{-h})s + y) + |\rho| w\Big)\, g(x,y,w,s;\boldsymbol{\theta})\, \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\le \big(|x| + |\mu|\big) \iiint_{\mathbb{D}} \frac{1}{\sqrt{(1-e^{-h})s}}\, f_{Y,W,S}(y,w,s)\ \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\qquad + \iiint_{\mathbb{D}} |\beta|\, \sqrt{(1-e^{-h})s}\ f_{Y,W,S}(y,w,s)\ \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\qquad + \iiint_{\mathbb{D}} \left(|\beta|\sqrt{y} \;+\; \frac{|\rho|}{\sqrt{1-e^{-h}}}\, \frac{w}{\sqrt{s}}\right) f_{Y,W,S}(y,w,s)\ \mathrm{d}y\ \mathrm{d}w\ \mathrm{d}s \\
&\le \frac{|x| + |\mu|}{\sqrt{1-e^{-h}}}\, E[S^{-1/2}] \;+\; \sqrt{\beta^2\,(1-e^{-h})}\, E[S^{1/2}] \\
&\qquad + |\beta| E[W^{1/2}] \;+\; \frac{|\rho|}{\sqrt{1-e^{-h}}} E[S^{-1/2}] \cdot E[W] \qquad \text{(B.37)} \\
&\triangleq K_2(\boldsymbol{\theta})
\end{aligned}
$$

and

$$\begin{aligned}
\Big|\,\partial_5 m_{\boldsymbol{\theta}}(x)\Big| \;\le\; & \iiint_{\mathbb{D}} \Big|\,\partial_5\, g(x,y,w,s;\boldsymbol{\theta})\Big|\,\mathrm{d}y\;\mathrm{d}w\;\mathrm{d}s \\
\le \iiint_{\mathbb{D}} & \left(\frac{\nu}{\alpha}s + w + \frac{\nu}{h}\right) g(x,s,y,w;\boldsymbol{\theta})\,\mathrm{d}y\;\mathrm{d}w\;\mathrm{d}s \\
\le \frac{\nu}{\alpha\sqrt{1-e^{-h}}} & \,E[S^{1/2}] \,+\, \frac{1}{\sqrt{1-e^{-h}}}E[S^{-1/2}]\cdot E[W] \\
& + \frac{\nu}{h\sqrt{1-e^{-h}}}\,E[S^{-1/2}] \qquad\qquad \text{(B.38)} \\
\triangleq K_5(\boldsymbol{\theta}) &
\end{aligned}$$

Since we know

$$E[S^{-1/2}] = \frac{\Gamma(\nu-1/2)}{\sqrt{\alpha}\,\Gamma(\nu)}, \qquad E[S^{1/2}] = \frac{\sqrt{\alpha}\,\Gamma(\nu+1/2)}{\Gamma(\nu)}, \quad \text{and} \quad E[W] = \frac{\nu}{\alpha},$$

it is clear that

$$\int_{B_{a,2}} \iiint_{\mathbb{D}} \Big|\,\partial_2\, g(x,y,w,s;\boldsymbol{\theta})\Big|\,\mathrm{d}y\;\mathrm{d}w\;\mathrm{d}s\;\;\mathrm{d}\beta \;<\; \infty$$

and

$$\int_{B_{a,5}} \iiint_{\mathbb{D}} \Big|\,\partial_5\, g(x,y,w,s;\boldsymbol{\theta})\Big|\,\mathrm{d}y\;\mathrm{d}w\;\mathrm{d}s\;\;\mathrm{d}\alpha \;<\; \infty$$

So Lemma B.2.3 applies and one has

$$\partial_2 m_{\boldsymbol{\theta}}(x) \;=\; \iiint_{\mathbb{D}} \Big(x-\mu-\beta((1-e^{-h})s+y)-\rho w\Big)\, g(x,y,w,s;\boldsymbol{\theta})\;\mathrm{d}y\;\mathrm{d}w\;\mathrm{d}s \tag{B.39}$$

and

$$\partial_5 m_{\boldsymbol{\theta}}(x) \;=\; \iiint_{\mathbb{D}} \Big(-\frac{\nu}{\alpha}s - w + \frac{\nu}{h}\Big)\, g(x,y,w,s;\boldsymbol{\theta})\; \mathrm{d}y\; \mathrm{d}w\; \mathrm{d}s \tag{B.40}$$

The same approach can be used again to study the $\partial_{25} m_{\boldsymbol{\theta}}(x)$ with more involved computations and gets the following result:

$$\begin{aligned}\partial_{25} m_{\boldsymbol{\theta}}(x) = \iiint_{\mathbb{D}} &\Big(x - \mu - \beta((1-e^{-h})s + y) - \rho w\Big)\cdot \\ &\Big(-\frac{\nu}{\alpha}s - w + \frac{\nu}{h}\Big)\, g(x,y,w,s;\boldsymbol{\theta})\; \mathrm{d}y\; \mathrm{d}w\; \mathrm{d}s\end{aligned} \tag{B.41}$$

Remark These results are exactly the conclusion of Proposition B.2.4.

(2) We use Lemma B.2.12 to study the tail behavior of $\partial_2 m_{\boldsymbol{\theta}}(x)$, $\partial_5 m_{\boldsymbol{\theta}}(x)$ and $\partial_{25} m_{\boldsymbol{\theta}}(x)$ for large $|x|$. Using the expressions of these three derivatives we derived in step (1), we immediately have

$$\begin{aligned} \partial_2 m_{\boldsymbol{\theta}}(x) &\le \tilde{\Psi}_2(h,\boldsymbol{\theta})\,(1+|x|^2) m_{\boldsymbol{\theta}}(x) \\ \partial_5 m_{\boldsymbol{\theta}}(x) &\le \tilde{\Psi}_5(h,\boldsymbol{\theta})\,(1+|x|^2) m_{\boldsymbol{\theta}}(x) \\ \text{and} \qquad \partial_{25} m_{\boldsymbol{\theta}}(x) &\le \tilde{\Psi}_{25}(h,\boldsymbol{\theta})\,(1+|x|^3) m_{\boldsymbol{\theta}}(x) \end{aligned}$$

for sufficiently large $|x|$.

Now, we are ready to justify (B.35).

(ex.1). $\big(e^{-\delta^n} - 2\big) m_{\boldsymbol{\theta}}(x)$ is integrable with respect to x for fixed β and α.

(ex.2). For $\partial_{25}\left((e^{-\delta^n}-2)\, m_{\boldsymbol{\theta}}(x)\right)$ given by (B.36),

$$
\begin{aligned}
\left|\,\partial_{25}\left((e^{-\delta^n}-2)\, m_{\boldsymbol{\theta}}(x)\right)\right| &\le \frac{2}{m_{\boldsymbol{\theta}}(x)} \cdot \left|\,\partial_2 m_{\boldsymbol{\theta}}(x)\right| \cdot \left|\,\partial_5 m_{\boldsymbol{\theta}}(x)\right| \,+\, 4\left|\,\partial_{25} m_{\boldsymbol{\theta}}(x)\right| \\
&\le \frac{2}{m_{\boldsymbol{\theta}}(x)} \cdot K_2(\boldsymbol{\theta}) \cdot K_5(\boldsymbol{\theta}) \,+\, 4K_{25}(\boldsymbol{\theta})
\end{aligned}
$$

Since $m_{\boldsymbol{\theta}}(x)$ is bounded away from 0 (see its definition at (B.18)), $\frac{1}{m_{\boldsymbol{\theta}}(x)}$ is a bounded function for all finite x. Therefore, $(e^{-\delta^n}-2)\, m_{\boldsymbol{\theta}}(x)$ is absolutely continuous with respect to both β and α.

(ex.3). Notice that,

$$
\begin{aligned}
&\int_{\mathbb{R}} \left|\,\partial_{25}\left((e^{-\delta^n}-2)\, m_{\boldsymbol{\theta}}(x)\right)\right| \mathrm{d}x \\
&\le 2\int_{\mathbb{R}} \left|\frac{\partial_2 m_{\boldsymbol{\theta}}(x)}{m_{\boldsymbol{\theta}}(x)}\right| \cdot \left|\frac{\partial_5 m_{\boldsymbol{\theta}}(x)}{m_{\boldsymbol{\theta}}(x)}\right| \cdot m_{\boldsymbol{\theta}}(x)\, \mathrm{d}x \,+\, 4\int_{\mathbb{R}} \left|\,\partial_{25} m_{\boldsymbol{\theta}}(x)\right| \mathrm{d}x \\
&\le 2\,\tilde{\Psi}_2(h,\boldsymbol{\theta})\,\tilde{\Psi}_5(h,\boldsymbol{\theta}) \int_{\mathbb{R}} (1+|x|^2)^2 \cdot m_{\boldsymbol{\theta}}(x)\, \mathrm{d}x \\
&\qquad +\, 4\,\tilde{\Psi}_{25}(h,\boldsymbol{\theta}) \int_{\mathbb{R}} (1+|x|^3) \cdot m_{\boldsymbol{\theta}}(x)\, \mathrm{d}x
\end{aligned}
$$

It is not difficult to see the last line is a continuous function of the β and α, therefore, it is locally integrable in $\boldsymbol{B_a}$ with respect to β and α.

At last, Lemma B.2.3 applies and we have (B.35) verified. Using similar steps, one can justify the Taylor expansion in Lemma 3.2.3 is valid provided that $\nu > 7/2$

B.3 Deriving Asymptotic Normality by the Functional Delta Method

In Section 3.2, we discuss how to derive the consistency and asymptotic normality for the NEDE. Recall the key step is to show (3.17)

$$n^{1/2}\int \Big[A(\delta_n) - A(\delta_s^*)\Big]\nabla m_{\boldsymbol{\theta}}^*(x)\,\mathrm{d}x \to MVN(0, V).$$

In the coming short paragraph, we will present a different approach to study its asymptotic normality. To simplify the discussion, we shall shift the focus temporarily to the conventional minimum distance estimator where the kernel $k(x; t, h)$ is only used to smooth the empirical distribution function but not the model CDF. We still use the Γ-OU BN-S process as the model process. Let $\hat{f}_n(x) = \frac{1}{n}\sum k(x; X_i, h)$ denote the kernel density estimate and let $s(x)$ denote the true stationary density of X_i. Similar to Section 3.2, given a family of model $\{m_{\boldsymbol{\theta}}(x)\}$ indexed by unknown parameter $\boldsymbol{\theta}$, we compute the estimator $\hat{\boldsymbol{\theta}}$ by

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}}\ \rho(\hat{f}_n, m_{\boldsymbol{\theta}}) = \underset{\boldsymbol{\theta}\in\Theta}{\operatorname{argmin}} \int G(\hat{f}_n(x), m_{\boldsymbol{\theta}}(x)) m_{\boldsymbol{\theta}}(x)\,\mathrm{d}x\,. \tag{B.42}$$

One easily notices that, unlike the MDE studied in Chapter 3, the above disparity $\rho(\cdot,\cdot)$ contains no kernel smoothed model density $m_{\boldsymbol{\theta}}(x)$. Basu et al. [14] studied the above estimate with i.i.d. data where $G(\cdot)$ corresponds to the Negative Exponential disparity. In what follows, we use the functional delta method and study the asymptotic normality of

$$\sqrt{n}\int_{\mathbb{R}}\left[A\Big(\frac{\hat{f}_n}{m_{\boldsymbol{\theta}^s}}-1\Big)-A\Big(\frac{s}{m_{\boldsymbol{\theta}^s}}-1\Big)\right]\nabla m_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x \tag{B.43}$$

for Negative Exponential disparity function $G(\cdot)$ with the data generated by a β-mixing process. Notice that in (B.43), the integral is an integrated functional of the kernel density estimate $\hat{f}_n(x)$, and the results in Aït-Sahalia [1, 2] can be applied to derive its asymptotic distribution.

Remark The motivation for including this section is twofold. The delta method itself is an intuitive yet powerful method to study the limiting distribution of functions of random variables. The functional delta method and the associated Von Mises calculus are particularly useful for many M-estimators problems. In fact, we have used the conventional delta method in Chapter 4 to derive the asymptotic normality of the MOM estimators. Second, one will find the method to be shown cannot be used directly to study the MDE described in Section 3.2. We hope to use this section to motivate extending the functional delta method to wider class of statistical functionals. As the focus here is to present the functional delta method, we will only study the asymptotic distribution of (B.43) but not pursue the asymptotic normality for $\hat{\boldsymbol{\theta}}$. Consistency of $\hat{\boldsymbol{\theta}}$ can be proved by similar steps described in Section 3.1 and Section 3.2, we won't elaborate the details here.

For completeness of our discussion, we summarize the results by Aït-Sahalia in [1] and [2] in the following exposition. Consider $\mathbb{R}^d$-valued random variables

$X_1,\ X_2,\ldots,X_n$ identically distributed as $s(\cdot)$ with cumulative distribution function $S(\mathbf{x}) = \int_{-\infty}^{\mathbf{x}} s(\mathbf{t})\,\mathrm{d}\mathbf{t}$ where $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. Assume the following regularity conditions:

(F-D1). The sequence $\{X_i\}$ is a strictly stationary β-mixing sequence satisfying:

$$k^{\Delta}\beta_k \to 0$$

for some fixed $\Delta > 1$ as $k \to \infty$.

(F-D2). The density function $s(\cdot)$ is continuously differentiable on $\mathbb{R}^d$ up to order s. Its successive derivatives are bounded and in $L_2(\mathbb{R}^d)$. Denote C^s as the space of density functions satisfying this assumption.

(F-D3). For the kernel K used to compute $\hat{f}_n(x)$, assume

(i) K is an even function integrating to one;

(ii) The kernel is of order $r = s$ where r is an even interger such that:

1) $\forall p \in \mathbb{N}^d$ with $|p| \equiv p_1 + \ldots + p_d \in \{1, \ldots, r-1\}$, one has

$$\int_{-\infty}^{\infty} x_1^{p_1} \cdots x_d^{p_d}\ K(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0$$

2) $\exists p \in \mathbb{N}^d$ with $|p| = r$ and

$$\int_{-\infty}^{\infty} x_1^{p_1} \cdots x_d^{p_d}\ K(\mathbf{x})\,\mathrm{d}\mathbf{x} \neq 0$$

3)

$$\int_{-\infty}^{\infty} \|\mathbf{x}\|^r \left|K(\mathbf{x})\right| \mathrm{d}\mathbf{x} < \infty$$

(iii) K is continuously differentiable up to order $s+d$ on $\mathbb{R}^d$, and its derivatives of order up to s are in $L_2(\mathbb{R}^d)$.

(F-D4). As $n \to \infty$, the bandwidth $h_n \to 0$ in such a rate that

$$n^{1/2}h_n^e + \left(n^{1/2}h_n^{2m}\right)^{-1} \to 0.$$

Assumption A4 is also denoted as A4(e,m).

Consider a functional $\Phi[.]$ defined on an open subset of C^s with the L_2 norm and taking values in $\mathbb{R}$. We say Φ is $L(2,m)$-differentiable at F in C^s if it admits a first order Taylor expansion:

$$\Phi[F+H] = \Phi[H] + \Phi^{(1)}[F](H) + R_{\Phi}[F+H]$$

with $R_{\Phi}[F+H] = O(\|H\|^2_{L(2,m)})$, where $\Phi^{(1)}[F](\cdot)$ is a continuous linear functional (in H) and $L(2,m)$ is the sum of the L_2 norm of all the derivatives of H up to order m. If the above expansion holds uniformly on H in any compact subset K of C^s and $|\Phi^{(1)}[F](H)| \leq C(K)\|H\|_{L(2,s)}$, then Φ is said to be $L(2,m)$-*Hadamard-differentiable* at F.

Remark For more discussion on differentiablity on statistical functionals, see [1], [29] and [79].

Next, introduce the real-valued integrated functional $\Phi(F)$ given by:

$$\Phi(F) \equiv \int_{-\infty}^{\infty} \omega(x)\Psi\big(x, F^{(1)}(x), F^{(2)}(x), \ldots, F^{(m)}(x)\big)\,\mathrm{d}x\,,$$

$\Phi(F)$ is defined on an open subset of C^s with the L_2 norm. Focusing on the case where x is univariate and let $\hat{F}_n$ be the kernel CDF estimator of $\{X_i\}$ i.e., $\hat{F}_n = \int_{-\infty}^{x} \hat{f}_n(t)\,\mathrm{d}t$, the following lemma holds.

Lemma B.3.1: **(Aït-Sahalia 1993 Corollary 1, Aït-Sahalia 1995)** Assume that $\omega(x)$ is $(m-1)$ times continuously differentiable and that Ψ is $\max(2, m)$-times continuously dfferentiable. Then under Assumptions A1 - A4(r, m):

(i) The functional Φ defined on an open subset U^s of C^s by:

$$\Phi(F) \equiv \int_{-\infty}^{\infty} \omega(x)\Psi\big(x, F^{(1)}(x), F^{(2)}(x), \ldots, F^{(m)}(x)\big)\,\mathrm{d}x$$

is $L(2, m)$-Hadamard-differentiable at the true CDF F with functional derivative given by:

$$\varphi[F](x) = \sum_{q=1}^{m} (-1)^{q-1} \frac{\partial^{q-1}}{\partial x^{q-1}}\Big(\omega(x)\frac{\partial \Psi}{\partial F^{(q)}}\big(x, F^{(1)}(x), \ldots, F^{(m)}(x)\big)\Big) \quad \text{(B.44)}$$

(ii)

$$\sqrt{n}\big\{\Phi(\hat{F}_n) - \Phi(S)\big\} \xrightarrow{\mathcal{D}} N(0, V_\Phi[F])$$

with:

$$V_{\Phi}[F] = \int_{-\infty}^{\infty} \varphi[F](x)^2 s(x)\,\mathrm{d}x - \Big(\int_{-\infty}^{\infty} \varphi[F](x) s(x)\,\mathrm{d}x\Big)^2 \qquad \text{(B.45)}$$
$$+ 2\sum_{k=1}^{\infty} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \big(s_k(x,y) - s(x)s(y)\big)\,\varphi[F](y)\,\varphi[F](x)\ \mathrm{d}y\ \mathrm{d}x$$

where $s_k(x, y)$ is the joint density of (X_i, X_{i+k}).

Remark Aït-Sahalia ([1] 1993) studied the functional delta method with assumption (F-D2) given by

The CDF S is continuously differentiable on $\mathbb{R}^d$ up to order $s + d$. The density $s(\cdot)$ has a compact support contained in $\mathbb{R}^d$. $s(\cdot)$ and its derivatives are zero on the boundary of the support.

Also, he used L_∞ norm to study the derivatives of Φ with respect to F. Whereas in [2], Aït-Sahalia relaxed the condition to allow for CDF with unbounded support and use L_2 norm to studied the functional derivative. Lemma B.3.1 above is stated in the form of [1] with the corresponding norm changed from $L(\infty, m)$ to $L(2, m)$.

Remark Assumption (F-D2) considers the regularity of the true underlying density, which also guarantees the finiteness of the variance of $\hat{f}_n$. Assumption (F-D3) is satisfied in our study as we use the Gaussian kernel. Assumption (F-D4) is a standard assumption on the bandwidth which makes sure the bias of the kernel density estimate goes to 0 as sample size increases.

To apply Lemma B.3.1 to study (B.43), we need to compute the derivative of the integrated functional

$$\Phi[S] \;=\; \int_{\mathbb{R}} A\Big(\frac{S^{(1)}}{m_{\boldsymbol{\theta}^s}} - 1\Big)\nabla m_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x\,.$$

It is easy to see that

$$\Psi\big(x, S^{(1)}(x), S^{(2)}(x), \ldots, S^{(m)}(x)\big) \;=\; A\Big(\frac{S^{(1)}}{m_{\boldsymbol{\theta}^s}} - 1\Big)$$

According to (B.44), $m = 1$ and $\frac{\partial \Psi}{\partial S^{(1)}} = A'\big(\delta_s\big) \cdot \frac{1}{m_{\boldsymbol{\theta}^s}(x)}$. Therefore, the derivative $\varphi[F](\cdot)$ of the functional $\Phi[S]$ is given by

$$\varphi[F](x) = A'\big(\delta_s\big) \cdot \nabla \ln m_{\boldsymbol{\theta}^s}(x) \tag{B.46}$$

By Lemma B.3.1,

Lemma B.3.2: Assume the conditions in Lemma B.3.1 hold and $A(\delta)$ is a regular RAF. If

$$V \;=\; E(\varphi[F](X_0)^{\otimes 2}) \;+\; 2\sum_{j=1}^{\infty} E(\varphi[F](X_0)\,\varphi[F](X_j)^T) \;<\; \infty,$$

then

$$\sqrt{n}\int_{\mathbb{R}} \Big[A\Big(\frac{\hat{f}_n}{m_{\boldsymbol{\theta}^s}} - 1\Big) - A\Big(\frac{s}{m_{\boldsymbol{\theta}^s}} - 1\Big)\Big]\nabla m_{\boldsymbol{\theta}^s}(x)\,\mathrm{d}x \;\xrightarrow{\mathcal{D}}\; MVN(0, V)$$

as $n \to \infty$.

C. MOMENTS AND CROSS-MOMENTS COMPUTATION

We begin this section by referring to a result in Cont and Tankov ([23] Sec 2.2.5) which considers the relation between moments, cumulants [1] and central moments. This result will help determining whether to use sample absolute moments or sample central moments when constructing the MOM equations.

Let X be a random variable and $\phi(u)$ be its characteristic function. If $\phi(u) \neq 0$ in a neighborhood of $u = 0$, then one can define a continuous function $\psi_X(u)$ as the logarithm of $\phi(u)$ in the neighborhood of zero by

$$\psi_X(0) = 0 \quad \text{and} \quad \phi_X(u) = \exp[\psi_X(u)]. \tag{C.1}$$

The $\psi_X(u)$ is called the *Cumulant Generating Function (CGF)* of X[2]. If $\phi(u) \neq 0$ for all u, then ψ_X can be extended to all $\mathbb{R}$. The k-th *Cumulant* is defined as

$$c_k(X) = \frac{1}{i^k} \frac{\partial^k \psi_X(u)}{\partial u^k}\bigg|_{u=0}.$$

We can define the k-th moment m_k for X similarly by

$$m_k = \frac{\partial^k M_X(u)}{\partial u^k}\bigg|_{u=0} \qquad \text{for } k \leq K$$

[1] In [23], the authors used the Cumulant Generating function to define cumulants. When the moment generating function is well defined, one can also use the Cumulant Transform function defined in (1.2) to compute the cumulants.

[2] It is also called log-characteristic and characteristic exponent in different literatures

provided that the *Moment Generating Function (MGF)* $M_X(u) = E[\exp(uX)]$ and its first K derivatives are well defined in the neighborhood of 0.

Denote the k-th central moment of X by $\mu_k(X) = E[(X - E\,X)^k]$, then $c_k(X)$, $\mu_k(X)$ and $m_k(X)$ for $k = 1, 2, 3, 4$ are related in the following way:

$$
\begin{aligned}
c_1(X) &= m_1(X) &&= E\,X, \\
c_2(X) &= \mu_2(X) &&= m_2(X) - m_1(X)^2 = \mathrm{Var}(X), \\
c_3(X) &= \mu_3(X) &&= m_3(X) - 3m_2(X)m_1(X) + 2m_1(X)^3, \\
c_4(X) &= \mu_4(X) &&- 3\mu_2(X)
\end{aligned}
$$

See [23] Section 2.2 for more details.

C.1 Moments of X_1

Recall we derive $\phi_X(u)$ in Chapter 2 with the assumption that $\lambda = 1$. In this section, we will drop this assumption when using $\phi_X(u)$ to compute the moments of X_1.

Redefine the functions g_1 and g_2 as follows:

$$
\begin{aligned}
g_1(s; u, \lambda, \beta) &= -\tfrac{(1-e^{-\lambda h})e^{-s}}{2\lambda}u^2 + i\beta\tfrac{(1-e^{-h})e^{-s}}{\lambda}u\,, \\
g_2(s; u, \lambda, \beta, \rho) &= -\tfrac{1-e^{-\lambda h+\lambda s}}{2\lambda}u^2 + i\rho u + i\,\tfrac{1-e^{-\lambda h+\lambda s}}{\lambda}\beta u\,.
\end{aligned}
$$

The characteristic function of X_1 is given by:

$$\begin{aligned}
\phi_{X_1}(u) &= E\left[e^{iu\mu h}\, e^{iu\beta\int_0^h \sigma_s^2 ds}\, e^{iu\rho\int_0^h dZ_s}\, e^{-\frac{u^2}{2}\int_0^h \sigma_s^2 ds}\right] \\
&= e^{iu\mu h} \exp\Big(\int_0^\infty \int_{\mathbb{R}+}(e^{g_1 x}-1)w(x)\ \mathrm{d}x\ ds\Big) \\
&\qquad \exp\Big(\lambda\int_0^h \int_{\mathbb{R}+}(e^{g_2 x}-1)w(x)\ \mathrm{d}x\ ds\Big) \\
&= e^{iu\mu h}\, \phi_1(u;\beta,\nu,\alpha)\, \phi_2(u;\beta,\rho,\nu,\alpha)
\end{aligned}$$

It turns out that even in the Γ-OU case where $w(x)$ takes the simple form $\nu\alpha e^{-\alpha x}$, $\phi_2(u)$ can end up to be very complicated. Therefore, we will only include the complete expression for $\phi_2(u)$ in the Γ-OU case for illustration.

1 Γ-OU BN-S Model. In this case, $\kappa(\theta) = \frac{\nu\theta}{\alpha-\theta}$ and $w(x) = \nu\alpha e^{-\alpha x}$ for the Compound Poisson process Z_t. After some computations, one can get

$$\phi_1(u;\beta,\nu,\alpha) = \left(1 + \frac{(1-e^{-\lambda h})u^2 - 2i\beta(1-e^{-\lambda h})u}{2\alpha\lambda}\right)^{-\nu}.$$

Define the following functions:

$$\begin{aligned}
f_1(u;\lambda,\beta,\rho,\nu,\alpha) &= \frac{\alpha\nu}{\lambda(u^2+2\alpha-2iu(\beta+\rho))} \\
f_2(u;\beta,\rho,\alpha) &= 2u\alpha\beta - u^3\rho \\
f_3(u;\lambda,\beta,\rho,\nu,\alpha) &= u^2(\alpha+2\rho(\beta+\rho)) + 2\alpha^2
\end{aligned}$$

$$\begin{aligned}
f_4(u;\lambda,\beta,\rho,\nu,\alpha) = (e^h-1)u^4 \ &+\ 4e^h\alpha^2 \\
&+\ \Big((4e^h-2)\alpha + 4(\beta+\rho)\big((e^h-1)\beta + e^h\rho\big)\Big)u^2
\end{aligned}$$

$$f_5(u;\lambda,\beta,\rho,\nu,\alpha) = (e^h-1)^2u^4 \; + \; 4e^{2h}\alpha^2 \\ + \; 4\Big(\beta^2 - e^h(\alpha+2\beta(\beta+\rho)) + e^{2h}\big((\beta+\rho)^2+\alpha\big)\Big)u^2$$

and $\phi_2(u;\beta,\rho,\nu,\alpha)$ is given by

$$\phi_2(u;\beta,\rho,\nu,\alpha) = \exp\Big\{-\lambda h\nu + f_1\Big(2h - \log\big(4e^{2h}(\alpha^2+u^2\rho^2)\big) + \log(f_5) \\ - 2i\,\arctan\big[\frac{f_2}{f_3}\big] - 2i\,\arctan\big[-2\frac{f_2}{f_4}\big]\Big)\Big\}$$

The mean and variance of X_1 are given by:

$$E[X_1] = h\mu + \frac{h\nu(\beta+\lambda\rho)}{\alpha} \tag{C.2}$$
$$\mathrm{Var}[X_1] = \frac{\nu}{\alpha^2\lambda^2}\Big((2\beta^2+4\beta\lambda\rho)(e^{-\lambda h}+(\lambda h-1)) + h\lambda^2(\alpha+2\lambda\rho^2)\Big)$$

2 IG-OU BN-S Model. In the IG-OU case, $\kappa(\theta) = \theta\delta(\gamma^2-2\theta)^{-1/2}$ and $w(x) = \frac{\delta}{2\sqrt{2\pi}}x^{-3/2}(1+\gamma^2x)e^{-\frac{1}{2}\gamma^2x}$ for the BDLP Z_t. $\phi_1(u;\beta,\delta,\gamma))$ is given by:

$$\phi_1(u,\beta,\delta,\gamma) = \delta\Big(\gamma - \sqrt{\gamma^2+(1-e^{-h})u(u-2i\beta)}\Big).$$

We have the mean and variance of X_1:

$$E[X_1] = \frac{h(\gamma\mu+\beta\delta+\delta\lambda\rho)}{\gamma} \tag{C.3}$$
$$\mathrm{Var}[X_1] = \frac{\delta}{\gamma^3\lambda^2}\Big((2\beta^2+4\beta\lambda\rho)(e^{-\lambda h}+(\lambda h-1)) + h\lambda^2\big(\gamma^2+2\lambda\rho^2\big)\Big)$$

C.2 Covariance of $(R_{t+h} - R_t, \sigma^2_{t+h} - \sigma^2_t)$, $(R_h - R_0, \sigma^2_h)$ and (X_j, X_k)

1. Covariance of $(R_{t+h} - R_t, \sigma^2_{t+h} - \sigma^2_t)$

By the definition of the BN-S model (1.1),

$$R_{t+h} - R_t \;=\; \mu h \;+\; \beta \int_t^{t+h} \sigma_s^2 \,\mathrm{d}s \;+\; \int_t^{t+h} \sigma_s \,\mathrm{d}W_s \;+\; \rho \int_t^{t+h} \mathrm{d}Z_{\lambda s}\,.$$

We compute the covariance between $R_{t+h} - R_t$ and $\sigma^2_{t+h} - \sigma^2_t$ to demonstrate that parameter ρ does control whether the increments of R_t and σ^2_t are positively or negatively correlated.

$$\begin{aligned}
&\mathrm{Cov}[R_{t+h} - R_t, \sigma^2_{t+h} - \sigma^2_t] \\
&\quad= E[(R_{t+h} - R_t)(\sigma^2_{t+h} - \sigma^2_t)] - E[(\sigma^2_{t+h} - \sigma^2_t)] \cdot E[(R_{t+h} - R_t)] \\
&\quad= E[(\sigma^2_{t+h} - \sigma^2_t)\, \mu h] \;+\; \beta\, E[\textstyle\int_t^{t+h} (\sigma^2_{t+h}\sigma^2_s - \sigma^2_t\sigma^2_s) ds] \\
&\qquad\quad + E[(\sigma^2_{t+h} - \sigma^2_t) \textstyle\int_t^{t+h} \sigma_s d\,W_s] \;+\; \rho E[(\sigma^2_{t+h} - \sigma^2_t) \textstyle\int_t^{t+h} \mathrm{d}Z_{\lambda s}\,] \\
&\quad= \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}
\end{aligned}$$

Second equality holds due to the stationarity of σ^2_t. For these four terms we have:

$\mathrm{I} \;=\; 0$ since σ^2_t is stationary;

$\mathrm{III} \;=\; 0$ since W_t is a standard Brownian motion and it's independent of σ^2_t;

When $t \leq s \leq t+h$, one has:

$$
\begin{aligned}
E[\sigma_t^2 \sigma_s^2] &= \mathrm{Cov}(\sigma_t^2, \sigma_s^2) + E[\sigma_t^2] \cdot E[\sigma_s^2] \\
&= \mathrm{Corr}(\sigma_t^2, \sigma_s^2)\,\mathrm{Var}(\sigma_0^2) + E(\sigma_0^2)^2 \\
&= e^{-\lambda(s-t)} \cdot \mathrm{Var}(\sigma_0^2) + E(\sigma_0^2)^2 \quad \text{and} \\
E[\sigma_{t+h}^2 \sigma_s^2] &= e^{-\lambda(t+h-s)} \cdot \mathrm{Var}(\sigma_0^2) + E(\sigma_0^2)^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{II} &= \beta \int_t^{t+h} e^{-\lambda(t+h-s)} ds \cdot \mathrm{Var}(\sigma_0^2) - \tfrac{1}{2} \int_t^{t+h} e^{-\lambda(s-t)} ds \cdot \mathrm{Var}(\sigma_0^2) \\
&= \Big(\beta\, [-\tfrac{1}{\lambda}\, e^{-\lambda(t+h-s)}\big|_t^{t+h}] - \beta\, [-\tfrac{1}{\lambda}\, e^{-\lambda(s-t)}\big|_t^{t+h}]\Big) \cdot \mathrm{Var}(\sigma_0^2) \\
&= 0
\end{aligned}
$$

For the last term:

$$
\begin{aligned}
\mathrm{IV} &= \rho E[(e^{-\lambda h} - 1)\sigma_t^2 \int_t^{t+h} \mathrm{d}Z_{\lambda s} + \int_t^{t+h} e^{-\lambda(t+h-u)} dZ_{\lambda u} \cdot \int_t^{t+h} dZ_{\lambda u}] \\
&= \mathrm{IV.1} + \mathrm{IV.2}
\end{aligned}
$$

Due to the independent increment properties of Lévy process Z_t, σ_t^2 is independent of $\int_t^{t+h} \mathrm{d}Z_{\lambda s}$. We get

$$
\mathrm{IV.1} = \rho\lambda h(e^{-\lambda h} - 1)E[\sigma_0^2]E[Z_1] = \rho\lambda h(e^{-\lambda h} - 1)E[Z_1]^2.
$$

To compute IV.2, first notice that Z_t is a subordinator, so it is of finite variation and the stochastic integral can be understood in the Lebesgue-Stieljes sense. Thus for any refining partition[3] $\pi_n = \{t = T_0 < T_1 < T_2 < \ldots < T_n = t+h\}$ whose grid size converges to 0 as $n \to \infty$,

[3]A sequence of partitions $\{\pi_n\}$ is called *refining* if the set of partition points $\{T_j^m\}$ is a subset of $\{T_j^n\}$ for all $m < n$

$$\textstyle\sum_{i=0}^{n-1}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i}) \quad \xrightarrow{a.s.} \quad \int_t^{t+h} dZ_{\lambda u} \qquad \text{and}$$

$$\textstyle\sum_{i=0}^{n-1} e^{\lambda T_i}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i}) \quad \xrightarrow{a.s.} \quad \int_t^{t+h} e^{\lambda u} dZ_{\lambda u}$$

This implies

$$\begin{aligned} I_n \quad &\triangleq \quad \textstyle\sum_{i=0}^{n-1} e^{\lambda T_i}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i}) \cdot \sum_{i=0}^{n-1}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i}) \\ &\xrightarrow{a.s.} I \triangleq \textstyle\int_t^{t+h} e^{\lambda u}\, \mathrm{d}Z_{\lambda u} \cdot \int_t^{t+h} \mathrm{d}Z_{\lambda u} \; = \; e^{-\lambda(t+h)} \, \text{IV.2} \end{aligned}$$

as $n \to \infty$. Next consider the random variable $V \triangleq e^{\lambda(t+h)}(Z_{\lambda(t+h)} - Z_{\lambda h})^2$. For any given partition π_n, using telescoping sum we can rewrite V as $\sum_{i=0}^{n-1} e^{\lambda(t+h)}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i}) \cdot \sum_{i=0}^{n-1}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i})$, from which we see $V \geq I_n$ a.s. since $e^{\lambda(t+h)} \geq e^{\lambda T_i}$. If we assume Z_1 have finite variance, then $E[V] = e^{\lambda(t+h)} E[Z_{\lambda h}^2] < \infty$, so $E[I_n] \to E[I]$ as $n \to \infty$ by the Dominated Convergence theorem. Let π_n denotes the equi-spaced partition with $T_i = t + \frac{ih}{n}$ for $n = 2k$, $k \in \mathbb{Z}^+$, We will use the limit of $E[I_n]$ to compute $E[I]$.

$$\begin{aligned} E[I_n] = \; & E\Big[\sum_{i=0}^{n-1} e^{\lambda T_i}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i}) \cdot \sum_{i=0}^{n-1}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i})\Big] \\ = \; & E\Big[\; \sum_{i=0}^{n-1} e^{\lambda T_i}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i})^2 + \\ & \quad \sum_{i=0}^{n-1}\sum_{j \neq i} e^{\lambda T_i}(Z_{\lambda T_{i+1}} - Z_{\lambda T_i})(Z_{\lambda T_{j+1}} - Z_{\lambda T_j}) \;\Big] \\ = \; & \sum_{i=0}^{n-1} e^{\lambda(t+ih/n)} E[Z_{\frac{\lambda h}{n}}^2] \; + \; \sum_{i=0}^{n-1}\sum_{j \neq i}^{n-1} e^{\lambda(t+ih/n)} E[Z_{\frac{\lambda h}{n}}]^2 \end{aligned}$$

Since

$$\begin{aligned} \sum_{i=0}^{n-1} E[Z_{\frac{\lambda h}{n}}^2] &= \sum_{i=0}^{n-1} \mathrm{Var}[Z_{\frac{\lambda h}{n}}] \; + \; \sum_{i=0}^{n-1} E[Z_{\frac{\lambda h}{n}}]^2 \\ &= \sum_{i=0}^{n-1} \frac{\lambda h}{n} \mathrm{Var}[Z_1] \; + \; n\,(\frac{\lambda h}{n})^2 E[Z_1]^2, \end{aligned}$$

$$
\begin{aligned}
E[I_n] &= e^{\lambda t}\left[\sum_{i=0}^{n-1} e^{i\frac{\lambda h}{n}}\frac{\lambda h}{n}\cdot \mathrm{Var}[Z_1] \;+\; \sum_{i=0}^{n-1} e^{i\frac{\lambda h}{n}}(\frac{\lambda h}{n})^2\cdot n\cdot E[Z_1]^2\right] \\
&= e^{\lambda t}\left[\lambda h\cdot \mathrm{Var}[Z_1]\frac{1}{n}\frac{e^{\lambda h}-1}{e^{\frac{\lambda h}{n}}-1} \;+\; \lambda^2 h^2\cdot E[Z_1]^2\frac{1}{n}\frac{e^{\lambda h}-1}{e^{\frac{\lambda h}{n}}-1}\right] \\
&\to e^{\lambda t}\left[\lambda h\cdot \mathrm{Var}[Z_1]\frac{e^{\lambda h}-1}{\lambda h} \;+\; \lambda^2 h^2\cdot E[Z_1]^2\frac{e^{\lambda h}-1}{\lambda h}\right] \\
&= e^{\lambda t}(e^{\lambda h}-1)\cdot \mathrm{Var}[Z_1] \;+\; e^{\lambda t}(e^{\lambda h}-1)\lambda h\cdot E[Z_1]^2 \\
&= E[I]
\end{aligned}
$$

From the above result we have IV.2 $= \rho(1-e^{-\lambda h})\big[\mathrm{Var}[Z_1] + \lambda h E[Z_1]^2\big]$, so

$$
\mathrm{Cov}(R_{t+h}-R_t, \sigma^2_{t+h}-\sigma^2_t) \;=\; \mathrm{IV} \;=\; \rho(1-e^{-\lambda h})\mathrm{Var}[Z_1]. \tag{C.4}
$$

2. Covariance of $(R_t - R_0, \sigma^2_h)$

The covariance to be computed will be used in Chapter 4 to construct the MOM estimators. Using the definition of R_t,

$$
\begin{aligned}
&\mathrm{Cov}(R_h - R_0, \sigma^2_h) \\
&= \;\; E[(R_h - R_0)\,\sigma^2_h] - E[\sigma^2_h]\cdot E[R_h - R_0] \\
&= \;\; E[\mu h\sigma^2_h] \;+\; \beta\, E[\textstyle\int_0^h \sigma^2_s\,\mathrm{d}s\;\sigma^2_h] \;+\; E[\textstyle\int_0^h \sigma_s\,\mathrm{d}W_s\,\sigma^2_h] \;+\; \rho\, E[\textstyle\int_0^h \mathrm{d}Z_{\lambda s}\;\sigma^2_h] \\
&\qquad -E[\sigma^2_h]\cdot\big[\mu + \beta h E[\sigma^2_h] \;+\; \rho\lambda h\, E[Z_1]\big] \\
&= \;\; \beta\, E[\textstyle\int_0^h \sigma^2_s\,\mathrm{d}s\;\sigma^2_h] \;+\; \rho\, E[\textstyle\int_0^h \mathrm{d}Z_{\lambda s}\;\sigma^2_h] \;-\; \beta h\big(E[\sigma^2_0]\big)^2 \;-\; \rho\lambda h\big(E[\sigma^2_0]\big)^2
\end{aligned}
$$

Recall term II in the computations of $\mathrm{Cov}(R_{t+h}-R_t, \sigma^2_{t+h}-\sigma^2_t)$,

$$
\begin{aligned}
E\Big[\int_0^h \sigma^2_h\,\sigma^2_s\,\mathrm{d}s\,\Big] &= \int_0^h \Big[e^{-\lambda(h-s)}\cdot \mathrm{Var}(\sigma^2_0) \;+\; \big(E[\sigma^2_0]\big)^2\Big]\,\mathrm{d}s \\
&= \frac{1}{\lambda}\,\big(1-e^{-\lambda h}\big)\mathrm{Var}(\sigma^2_0) \;+\; h\big(E[\sigma^2_0]\big)^2
\end{aligned}
$$

and use term IV in $\mathrm{Cov}(R_{t+h} - R_t, \sigma^2_{t+h} - \sigma^2_t)$,

$$\begin{aligned}
\rho\, E\Big[\int_0^h \mathrm{d}Z_{\lambda s}\ \sigma_h^2\Big] &= \rho\,\lambda h\, e^{-\lambda h}\,\big(E[\sigma_0^2]\big)^2 \;+\; \rho\big(1-e^{-\lambda h}\big)\Big[\mathrm{Var}(Z_1) + \lambda h E[Z_1]^2\Big] \\
&= \rho\,\lambda h\, e^{-\lambda h}\,\big(E[\sigma_0^2]\big)^2 \;+\; \rho\big(1-e^{-\lambda h}\big)\Big[2\,\mathrm{Var}(\sigma_0^2) + \;\lambda h\big(E[\sigma_0^2]\big)^2\Big] \\
&= 2\rho\big(1-e^{-\lambda h}\big)\cdot \mathrm{Var}(\sigma_0^2) \;+\; \rho\lambda h\big(E[\sigma_0^2]\big)^2
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{Cov}(R_h - R_0, \sigma_h^2) = \;& \frac{\beta\big(1-e^{-\lambda h}\big)}{\lambda}\,\mathrm{Var}(\sigma_0^2) \;+\; \beta h\big(E[\sigma_0^2]\big)^2 \\
&+\; 2\rho\big(1-e^{-\lambda h}\big)\,\mathrm{Var}(\sigma_0^2) \;+\; \rho\lambda h\big(E[\sigma_0^2]\big)^2 \\
&-\; \beta h\big(E[\sigma_0^2]\big)^2 \;-\; \rho\lambda h\big(E[\sigma_0^2]\big)^2 \\
=\;& \Big(\frac{\beta}{\lambda} + 2\rho\Big)\,\big(1-e^{-\lambda h}\big)\,\mathrm{Var}(\sigma_0^2). \qquad \text{(C.5)}
\end{aligned}$$

3. Covariance of (X_j, X_k)

Since X_i is strictly stationary, the covariance of (X_j, X_k) should only depends on $|j - k|$. This means it is sufficient to compute $\mathrm{Cov}(X_1, X_j)$ for $j > 1$. We will use the characteristic function of (X_1, X_j) to find their covariance. Recall if $\mathrm{E}[X_1\, X_j] < \infty$, then

$$\frac{1}{i^2}\frac{\partial^2}{\partial u_1\,\partial u_2}\phi_{X_1,X_j}(u_1,u_2)\Big|_{u_1=0,u_2=0} = \mathrm{E}[X_1\, X_j].$$

From the BN-S structure, we know

$$
\begin{aligned}
X_1 &= \mu h + \beta \int_0^h \sigma_s^2 \,\mathrm{d}s \ + \int_0^h \sigma_s dW_s + \rho \int_0^h \,\mathrm{d}Z_{\lambda s} \\
X_j &= \mu h + \beta \int_{(j-1)h}^{jh} \sigma_s^2 \,\mathrm{d}s \ + \int_{(j-1)h}^{jh} \sigma_s dW_s + \rho \int_{(j-1)h}^{jh} \,\mathrm{d}Z_{\lambda s}
\end{aligned}
$$

Since $\sigma_s^2 = e^{-\lambda s}\sigma_0^2 + e^{-\lambda s}\int_0^s e^{\lambda u}\,\mathrm{d}Z_{\lambda u}$, one has

$$
\begin{aligned}
\int_{(j-1)h}^{jh} \sigma_s^2 \,\mathrm{d}s \ &= \int_{(j-1)h}^{jh} e^{-\lambda s}\,\mathrm{d}s\,\sigma_0^2 + \int_{(j-1)h}^{jh}\int_0^s e^{\lambda u}\,\mathrm{d}Z_{\lambda u}\ e^{-\lambda s}\,\mathrm{d}s \\
&= \frac{1}{\lambda}\Big(e^{-(j-1)\lambda h} - e^{-j\lambda h}\Big)\sigma_0^2 \\
&\quad + \int_0^{(j-1)h}\int_{(j-1)h}^{jh} e^{-\lambda s}\,\mathrm{d}s\ e^{-\lambda u}\,\mathrm{d}Z_{\lambda u} \ + \int_{(j-1)h}^{jh}\int_u^{jh} e^{-\lambda s}\,\mathrm{d}s\ e^{-\lambda u}\,\mathrm{d}Z_{\lambda u} \\
&= \frac{1}{\lambda}\Big(e^{-(j-1)\lambda h} - e^{-j\lambda h}\Big)\sigma_0^2 \\
&\quad + \int_0^{(j-1)h} \frac{e^{\lambda u}}{\lambda}\Big(e^{-(j-1)\lambda h} - e^{-j\lambda h}\Big)\,\mathrm{d}Z_{\lambda u} \ + \int_{(j-1)h}^{jh} \frac{1 - e^{-j\lambda h + \lambda u}}{\lambda}\,\mathrm{d}Z_{\lambda u}
\end{aligned} \tag{C.6}
$$

To get the joint characteristic function for (X_1, X_j), let us first exam $iu_1X_1 + iu_2X_j$:

$$
\begin{aligned}
iu_1X_1 = i\mu h u_1 &+ \Big(i\beta u_1 - \frac{u_1^2}{2}\Big)\frac{1 - e^{-\lambda h}}{\lambda}\sigma_0^2 \\
&+ \int_0^h \Big(i\big(\rho + \beta\frac{1 - e^{-\lambda h + \lambda s}}{\lambda}\big)u_1 - \frac{1 - e^{-\lambda h + \lambda s}}{2\lambda}u_1^2\Big)\,\mathrm{d}Z_{\lambda s}
\end{aligned}
$$

$$
\begin{aligned}
iu_2X_j = i\mu h u_2 &+ \Big(i\beta u_2 - \frac{u_2^2}{2}\Big)\frac{e^{-(j-1)\lambda h} - e^{-j\lambda h}}{\lambda}\sigma_0^2 \\
&+ \int_0^{(j-1)h} \Big(\big(i\beta u_2 - \frac{u_2^2}{2}\big)\big(e^{-(j-1)\lambda h} - e^{-j\lambda h}\big)\frac{e^{\lambda s}}{\lambda}\Big)\,\mathrm{d}Z_{\lambda s} \\
&+ \int_{(j-1)h}^{jh} \Big(\frac{1}{\lambda}\big(i\beta u_2 - \frac{u_2^2}{2}\big)\big(1 - e^{-j\lambda h + \lambda s}\big) + i\rho u_2\Big)\,\mathrm{d}Z_{\lambda s}
\end{aligned}
$$

Using the fact that Z_t has independent increments over non-overlapped intervals, we can decompose the integrals in iu_2X_j into integrals over $(0,h]$, $(h,(j-1)h]$ and $((j-1)h, jh]$. Define the following functions:

$$
\begin{aligned}
g_1(u_1,u_2;s,\lambda,h,\beta) &= e^{-s}\left[\left(i\beta u_1 - \frac{u_1^2}{2}\right)\frac{1-e^{-\lambda h}}{\lambda} + \left(i\beta u_2 - \frac{u_2^2}{2}\right)\frac{e^{-(j-1)\lambda h} - e^{-j\lambda h}}{\lambda}\right] \\
g_2(u_1,u_2;s,\lambda,h,\beta,\rho) &= i\left(\rho + \beta\frac{1-e^{-\lambda h+\lambda s}}{\lambda}\right)u_1 - \frac{1-e^{-\lambda h+\lambda s}}{2\lambda}u_1^2 \\
&\quad + \left(i\beta u_2 - \frac{u_2^2}{2}\right)\left(e^{-(j-1)\lambda h} - e^{-j\lambda h}\right)\frac{e^{\lambda s}}{\lambda} \\
g_3(u_2;s,\lambda,h,\beta) &= \left(i\beta u_2 - \frac{u_2^2}{2}\right)\left(e^{-(j-1)\lambda h} - e^{-j\lambda h}\right)\frac{e^{\lambda s}}{\lambda} \\
g_4(u_2;s,\lambda,h,\beta,\rho) &= \frac{1}{\lambda}\left(i\beta u_2 - \frac{u_2^2}{2}\right)\left(1 - e^{-j\lambda h+\lambda s}\right) + i\rho u_2
\end{aligned}
$$

Then $\phi_{X_1,X_j}(u_1,u_2)$ is given by:

$$
\begin{aligned}
\phi_{X_1,X_j}(u_1,u_2) &= e^{i(u_1+u_2)\mu h}\cdot \mathrm{E}\Big[\exp\Big\{\int_0^\infty g_1 \mathrm{d}Z_s\Big\}\Big]\cdot \mathrm{E}\Big[\exp\Big\{\int_0^h g_2\, \mathrm{d}Z_{\lambda s}\Big\}\Big] \\
&\quad \cdot \mathrm{E}\Big[\exp\Big\{\int_h^{(j-1)h} g_3\, \mathrm{d}Z_{\lambda s}\Big\}\Big]\cdot \mathrm{E}\Big[\exp\Big\{\int_{(j-1)h}^{jh} g_4\, \mathrm{d}Z_{\lambda s}\Big\}\Big] \\
&= e^{i(u_1+u_2)\mu h}\cdot \exp\Big\{\int_0^\infty \kappa(g_1)ds\Big\}\cdot \exp\Big\{\lambda\int_0^h \kappa(g_2)ds\Big\} \\
&\quad \cdot \exp\Big\{\lambda\int_h^{(j-1)h} \kappa(g_3)ds\Big\}\cdot \exp\Big\{\lambda\int_{(j-1)h}^{jh} \kappa(g_4)ds\Big\} \qquad \text{(C.7)}
\end{aligned}
$$

where as before $\kappa(.)$ is the CTF of Z_t. To compute $\frac{1}{i^2}\frac{\partial^2}{\partial u_1\,\partial u_2}\phi_{X_1,X_j}(u_1,u_2)|_{u_1=0,u_2=0}$, one may use the following steps if one assumes $\int_0^\infty x^2 w(x)\,\mathrm{d}x < \infty$. Take $\phi_1 =$

$\exp\left(\int_0^\infty \kappa(g_1)ds\right)$ as example,

$$
\begin{aligned}
&\frac{\partial}{\partial u_1}\phi_1(u_1,u_2)\Big|_{u_1=0,u_2=0} \\
&= \phi_1(0,0)\cdot \lim_{\substack{u_1\to 0\\ u_2\to 0}} \frac{\partial}{\partial u_1}\int_0^\infty \int_{\mathbb{R}+}(e^{g_1(u_1,u_2;s,h,\beta,\lambda)x}-1)w(x)\,\mathrm{d}x\ \mathrm{d}s \\
&= \lim_{\substack{u_1\to 0\\ u_2\to 0}} \int_0^\infty \int_{\mathbb{R}+}\frac{\partial}{\partial u_1}(e^{g_1(u_1,u_2;s,h,\beta,\lambda)x}-1)w(x)\,\mathrm{d}x\ \mathrm{d}s \\
&= \lim_{\substack{u_1\to 0\\ u_2\to 0}} \int_0^\infty \int_{\mathbb{R}+}\left[e^{g_1(u_1,u_2;s,h,\beta,\lambda)x}\left(i\beta-u_1\right)\frac{1-e^{-\lambda h}}{\lambda}e^{-s}\right]x\,w(x)\,\mathrm{d}x\ \mathrm{d}s \\
&= \int_0^\infty \int_{\mathbb{R}+}\left[i\beta\frac{1-e^{-\lambda h}}{\lambda}e^{-s}\right]x\,w(x)\,\mathrm{d}x\ \mathrm{d}s
\end{aligned}
$$

One can verify the interchange differentiation and limit with integration under the assumption that $\int_0^\infty x^2 w(x)\,\mathrm{d}x < \infty$, thus the above simplification is valid. We can apply the above techniques to all ϕ_i's and compute the $\mathrm{E}[X_1\,X_j]$.

In the Γ-OU BN-S model:

$$
\begin{aligned}
\mathrm{Cov}(X_1,X_j) = {} & \frac{e^{-(j+1)\lambda h}\beta\nu}{\alpha^2\lambda^2}\left(e^{\lambda h}-1\right)^2 \\
& \Big(2\beta\nu + e^{\lambda h}\big(\beta(\nu\lambda h-\nu+1)+\lambda(\nu\lambda h+2)\rho\big)\Big)
\end{aligned}
\tag{C.8}
$$

In the IG-OU BN-S model:

$$
\begin{aligned}
\mathrm{Cov}(X_1,X_j) = {} & \frac{e^{-(j+1)\lambda h}\beta\delta}{\alpha^3\lambda^2}\left(e^{\lambda h}-1\right)^2 \\
& \Big(2\beta\delta\gamma + e^{\lambda h}\big(\beta(\delta\gamma(\lambda h-1)+1)+\lambda(\delta\gamma\lambda h+2)\rho\big)\Big)
\end{aligned}
\tag{C.9}
$$

Although both expressions are very involved, but it is not difficult to find that they both decay exponentially fast as $j \to \infty$. As a last remark, the approach we take to find the covariance between X_1 and X_j does not apply to finding the variance of X_1, because we have implicitly assumed $X_j \neq X_1$ when deriving their joint characteristic function.

BIBLIOGRAPHY

[1] Aït-Sahalia, Y. (1993). Nonparametric functional estimation with applications to financial models, *Ph.D. Thesis*, MIT.

[2] Aït-Sahalia, Y. (1995). The delta method for nonparametric kernel functionals, *Working Papar. Graduate School of Business, Univ. Chicago.*

[3] Bae, J. , Kim, C. and Nelson, C.R. (1997). Why are stock returns and volatility negatively correlated? *J. Empirical Finance*, Vol. **14** Issue 1, 41 - 58.

[4] Barndorff-Nielsen, O.E. (1997). Normal inverse Gaussian distributions and stochastic volatility modelling, *Scand. J. Stat.*, **24**(1), 1 - 13.

[5] Barndorff-Nielsen, O.E. (2001). Superposition of Ornstein-Uhlenbeck type processes, *Theor. Probab. Appl.*, **45**(2), 175 - 194.

[6] Barndorff-Nielsen, O.E. and Shephard, N. (2001a). Modelling by Lévy Processes for Financial Econometrics. In: O. E. Barndorff-Nielsen, T. Mikosch, S. Resnick (eds.), *Lévy Processes, Theory and Applications*, 283-318, Boston: Birkhäuser.

[7] Barndorff-Nielsen, O.E. and Shephard, N. (2001b). Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial econmetrics, *J. R. Stat. Soc. Ser. B Stat. Methodol.*, Vol. **63**, Issue 2, 167 - 241.

[8] Barndorff-Nielsen, O.E. and Shephard, N. (2002). Econometric analysis of realised volatility and its use in estimating stochastic volatility models. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, Vol. **6**, Issue 2, 253 - 280.

[9] Barndorff-Nielsen, O.E. and Shephard, N. (2003). Realised power variation and stochastic volatility, *Bernoulli*, Vol. **9** 243 - 265 (correction, p. 1109-1111).

[10] Barndorff-Nielsen, O.E. and Shephard, N. (2004). Power and bipower variation

with stochastic volatility and jumps (with discussion), *J. Financ. Economet.*, Vol. **2**, 1 - 48.

[11] Barndorff-Nielsen, O.E. , Shephard, N. and Winkel, M. (2006). Limit theorems for multipower variation in the presence of jumps, *Stochastic Process. Appl.*, Vol. **116**, Issue 5, 796 - 806.

[12] Basu, A. and Lindsay, B.G. (1994). Minimum disparity estimation for continuous models: efficiency, distributions and robustness, *Ann. Inst. Statist. Math.*, Vol. **46**, No.4, 683 - 705.

[13] Basu, A. and Sarkar, S. (1994). The trade-off between robustness and efficiency and the effect of model smoothing in minimum disparity inference, *J. Statist. Comput. Simulation*, Vol. **50**, Issue 3 & 4, 173 - 185.

[14] Basu, A., Sarkar, S. and Vidyashankar, A.N. (1994). Minimum negative exponential disparity estimation in parametric models, *J. Statist. Plann. Inference*, Vol. **58**, 349 - 370.

[15] Basu,A., Park,C., Lindsay,B.G. and Li, H. (2004). Some variants of Minimum Disparity estimation, *Comput. Statist. Data Anal.*, Vol. **45** 741 - 763.

[16] Beran, R.J. (1977). Minimum Hellinger distance estimates for parametric models, *Ann. Statist.*, Vol. **5**, 445 - 463

[17] Bertail, P. , Doukhan, P. and Soulier, P. (2006). *Dependence in Probability and Statistics (Lecture Notes in Statistics 187)*, Springer.

[18] Bhandari, S.K., Basu, A. and Sarkar, S. (2006). Robust inference in parametric models using the family of generalized negative exponential disparities, *Aust. N. Z. J. Stat.*, Vol. **48**, Issue 1, 95 - 114.

[19] Billingsley, P. (1995). *Probability and Measure, third edition* , Wiley.

[20] Broadie, M., Chernov, M. and Johannes, M. (2007). Model specification and risk premia: evidence from futures options, *J. Financ.*, Vol. **LXII**, No. 3, 1453 - 1490.

[21] Broto,C. and Ruiz, E. (2004). Estimation methods for stochastic volatility models: a survey, *J. Econ. Surv.*, Vol. **18**, Issue 5, 613 - 649.

[22] The CBOE Volatility Index - VIX (2009)

[23] Cont, R. and Tankov, P. (2004). *Financial Modeling With Jump Processes* , CRC Press.

[24] Donoho, D.L. and Liu, R.C. (1988). The "automatic" robustness of minimum distance functionals, *Ann. Statist.*, Vol. **16**, No. 2 552-586.

[25] Dotsis, G. , Raphael, N.M. and Mills, T.C. (2008). Estimation of continuous-time stochastic volatility models. In: *Handbook of Econometrics*, Vol. **2**, Palgrave Macmillan.

[26] Doukhan, P. (1994). *Mixing: Properties and Examples* (Lecture Notes in Statistics), Springer.

[27] Erdélyi, A. (1956). *Asymptotic Expansions*, Dover Publications.

[28] Fasen, V. (2010). Asymptotic results for the sample autocovariance function and extremems of integrated generalised Ornstein-Uhlenbeck processes, *Bernoulli*, Vol. **16**, No.1 51 - 79.

[29] Fernholz, L.T. (1983). *Von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics 19, Springer-Verlag.

[30] Fouque, J., Papanicolaou, G. and Sircar, K. R. (2000). *Derivatives in Finalcial Markets with Stochastic Volatility*, Cambridge Univ. Press.

[31] Ghysels, E., Harvey, A.C. and Renault, E. (1996). Stochastic volatility, In: *Handbook of Statistics*, Vol. **14**, G.S.Maddala and C.R.Rao, eds.

[32] Hampel, F.R., Ronchetti, E., Rousseeuw, P.J. and Stahel, W. (1986). *Robust Statistics: The Approach Based on Influence Functions*, Wiley, New York.

[33] Hansen, B. (2008). Uniform convergence rates for kernel estimation with dependent data, *Economet. Theor.* Vol.**24**, 726 - 748.

[34] Hansen, P. and Lunde, A. (2008). Realized variance and market microstructure noise, *J. Bus. Econom. Statist.*, Vol.**24**, 127 - 161.

[35] Haug, S. , Klüppelberg, C. , Lindner, A. and Zapp, M. (2007). Method of moment estimation in the COGARCH(1,1) Model, *Economet. J.* Vol.**10**, 320 - 341.

[36] Hubalek, F. and Posedel P. (2008). Joint analysis and estimation of stock prices and trading volume in Barndorff-Nielsen and Shephard stochastic volatility models, *Working Paper*.

[37] Hubalek, F. and Sgarra C. (2009). On the Esscher transforms and other equivalent martingale measures for Barndorff-Nielsen and Shephard stochastic volatility models with jumps, *Stoc. Proc. and App.* Vol.**119**, Issue 7, 2137 - 2157.

[38] Huber, P.J. and Ronchetti, E.M. (2009). *Robust Statistics, second edtion*, Wiley Series in Probability and Statistics.

[39] Ibragimov, I.A. and Linnik, Y.V. (1971). *Independent and Stationary Sequences of Random Variables*, Wolters-Noordhoff Publishing.

[40] Jacod, J. (1975). Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales, *Probab. Theory Related Fields*, Vol.**31**, No.3, 235 - 253.

[41] Jacod, J. and Shiryaev, A.N. (2003). *Limit Theorems For Stochastic Processes, second edition*, Springer.

[42] Jacod, J. (2008). Asymptotic properties of power variations and associated functionals of semimartingales, *Stochastic Process. Appl.*, Vol.**118**, 517 - 559.

[43] Kagan, A. and Nagaev S. (2001). How many moments can be estimated from a large sample? *Statist. Probab. Lett.* Vol.**55**, Issue 1, 99 - 105.

[44] Klüppelberg, C. ; Lindner, A. and Maller, R. (2004). A continous time GARCH process driven by a Lévy process: stationarity and second order behaviour, *Jour. Appl. Probab.* Vol. **41**, No.3, 601 - 622.

[45] Klüppelberg, C. ; Lindner, A. and Maller, R. (2006). Continuous time volatility modelling: COGARCH versus Ornstein-Uhlenbeck models, In: *From Stochastic Calculus to Mathematical Finance*, Kabanov, Y., Lipster, R. and Stoyanov, J. Eds., 393 - 419, Springer, Berlin.

[46] Kolmogorov, A.N. and Fomin S.V. (1975). *Introductory Real Analysis*, Translated & Edited by R.A.Silverman, Dover.

[47] Lehmann, E.L. and Casella, G. (1998). *Theory of Point Estimation, Second Edition*, Springer.

[48] Lin, Y-N (2007). Pricing VIX futures: evidence from integrated physical and risk-neutral probability measures, *J. Futures Markets* Vol.**27**, No.2, 1175 - 1217.

[49] Lindsay, B. G. (1994). Efficiency versus robustness: the case for minimum Hellinger distance and related methods, *Ann. Statist.*, Vol.**22**, No.2, 1081 - 1114.

[50] Lindner, A. and Maller, R.A. (2005). Levy integrals and the stationarity of generalised Ornstein-Uhlenbeck processes, *Stochastic Process. Appl.*, Vol.**115**, 1701 - 1722.

[51] Martin, R.D. and Yohai, V.J. (1986). Influence functionals for time series, *Ann. Statist.*, Vol.**14**, No.3, 781-818.

[52] Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process, *Bernoulli* **10**, No.1, 97 - 120.

[53] Meyn, Sean P. and Tweedie, R.L. (1992). Stability of Markovian processes I: criteria for discrete-time chains, *Adv. in Appl. Probab.*, Vol.**24**, No.3, 542 - 574

[54] Meyn, Sean P. and Tweedie, R.L. (2009). *Markov chains and stochastic stability, second edition*, Springer-Verlag.

[55] Nicolato, E. and Venardos, E. (2003). Option pricing in stochastic volatility models of the Ornstein-Uhlenbeck type, *Mathmatical Finance*, Vol.**13**, Issue 4, 445 - 466.

[56] Nummelin, E. and Tuominen, P. (1982). Geometric ergodicity of Harris recurrent Markov chains with application to renewal theory, *Stochastic Process. Appl.*, Vol.**12**, 187 - 202.

[57] Park, C. , Basu, A. and Lindsay, B.G. (2002). The residual adjustment function and weighted likelihood: a graphical interpretation of robustness of minimum disparity estiamtors, *Comput. Statist. Data Anal.*, Vol.**39**, 21 - 33.

[58] Rao, B.L.S. Prakasa (1999). *Semimartingales and their Statistical Inference*, Boca Raton : Chapman & Hall / CRC.

[59] Rao, C.R. (1961). Asymptotic efficiency and limiting information, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol.**1**, 531 - 546.

[60] Renault, E. (2009). Moment-based estimation of stochastic volatility models. In: *Handbook of Financial Time Series*, Springer.

[61] Protter, P.E.(2004). *Stochastic Integration and Differential Equations, second edition*, Springer, Berlin.

[62] Rudin, W. (1986). *Real and Complex Analysis*, McGRAW-Hill.

[63] Sato, K. (1999). *Lévy Processes and Infinitely Divisible Distributions*, Cambridge Univ. Press.

[64] Sato, K. and Yamazato, M. (1984). Operator-selfdecomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type, *Stochastic Process. Appl.*, Vol.**17**, 73 - 100.

[65] Simpson, D.G. (1987). Minimum Hellinger distance estimation for the analysis of count data, *J. Amer. Statist. Assoc.*, Vol.**82**, 802 - 807.

[66] Simpson, D.G. (1989). Hellinger deviance test: efficiency, breakdown points and examples, *J. Amer. Statist. Assoc.*, Vol.**84**, 107 - 113.

[67] Spiliopoulos, K (2008). Method of moments estimation of Ornstein-Uhlenbeck processes driven by general Lévy process, *In Preparation*

[68] Schoutens, W. (2003). *Lévy Processes in Finance: Pricing Financial Derivatives*, Wiley.

[69] Shao, J. (1999). *Mathematical Statistics*, Springer-Verlag.

[70] Takada, T. (2009). Simulated minimum Hellinger distance estimation of stochastic volatility models, *Comput. Statist. Data Anal.*, Vol.**53**, No.6, 2309 - 2403.

[71] Tamura, R.N. and Boos, D.D. (1986). Minimum Hellinger distance estimation for multivariate location and covariance, *J. Amer. Statist. Assoc.*, Vol.**81**, 223 - 229.

[72] Todorov, V. (2008). Volatility jumps, *Working Paper.*

[73] Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data, *J. Econometrics*, Vol.**148**, Issue 2, 131 - 148.

[74] Tran, L. (1989). The L1 convergence of kernel density estimates under dependence, *Canad. J. Statist.*, Vol.**17**, No.2, 197 - 208.

[75] Toronjadze, T. (2004). Stochastic equations in the problems of semimartingale parameter estimation, *J. Math. Sci. (N.Y.)*, Vol,**121** Issue 6, 2709-2840.

[76] Tuominen, P. and Tweedie, R.L. (1979). Exponential decay and ergodicity of general Markov processes and their discrete skeletons, *Adv. Appl. Probab.*, Vol.**11**, 784 - 803.

[77] Ushakov, N.G. (1999). *Selected Topics in Characteristic Functions*, Brill Academic Publishers.

[78] Veraart, A.E.D. (2010). Inference for the jump part of quadratic variation of Itô semimartingales, *Economet. Theor.*, Vol.**26**, 331 - 368.

[79] van der Vaart, A.W. (2000). *Asymptotic Statistics*, Cambridge Univ. Press.

[80] Woerner, J.H.C. (2003). Variational sums and power variation: a unifying approach to model selection and estimating in semimartingale models, *Statist. Decisions*, Vol.**21**, 47 - 68.

[81] Woerner, J.H.C. (2006). Power and multipower variation: inference for high frequency data. In: *Stochastic Finance*, Springer.