Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Estimation Results for the Generalized Langevin Equation with Lévy Jumps

por

Felipe Sousa Quintino

Brasília-DF
2021

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Estimation Results for the Generalized Langevin Equation with Lévy Jumps

por
Felipe Sousa Quintino[1]

Tese de Doutorado apresentada ao Programa de Pós-Graduação em Matemática do Departamento de Matemática da Universidade de Brasília, PPG-Mat–UnB, como parte dos requisitos necessários para obtenção do título de Doutor em Matemática, com habilitação no curso Matemática Aplicada com ênfase em Probabilidade.

sob orientação da

Profa. Dra. Chang Chung Yu Dorea

Brasília-DF
2021

[1]O autor contou com apoio financeiro CAPES e CNPq durante a realização deste trabalho.

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

QQ7e

Quintino, Felipe Sousa
Estimation Results for the Generalized Langevin Equation with Lévy Jumps / Felipe Sousa Quintino; orientador Chang Chung Yu Dorea. -- Brasília, 2021.
97 p.

Tese (Doutorado - Doutorado em Matemática) -- Universidade de Brasília, 2021.

1. Generalized Langevin Equation. 2. Generalized Ornstein-Uhlenbeck Processes. 3. Drift Estimation. 4. LAN Property. 5. Filtered MLE. I. Chung Yu Dorea, Chang , orient. II. Título.

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Matemática

Estimation Results for the Generalized Langevin Equation with Lévy Jumps

por

Felipe Sousa Quintino

Tese apresentada ao Departamento de Matemática da Universidade de Brasília, como parte dos requisitos para obtenção do grau de

DOUTOR EM MATEMÁTICA

Brasília, 25 de novembro de 2021.

Comissão Examinadora:

Profa. Dra. Chang Chung Yu Dorea- MAT/UnB (Orientadora)

Prof. Dr. Ary Vasconcelos Medino- MAT/UnB (Membro)

Prof. Dr. Hugo Alexander de la Cruz Cancino– FGV (Membro)

Profa. Dra. Verónica Andrea González López– UNICAMP (Membro)

Dedicatória

Dedico meu título de doutor à minha família que esteve ao meu lado todo esse tempo. Em especial à minha mãe que nunca deixou faltar papel e caneta para meus estudos.

Agradecimentos

Agradeço primeiramente à Deus por todas as oportunidades que tem me proporcionado até aqui. Agradeço também a minha família que sempre esteve ao meu lado nos momentos em que precisei. De forma especial à minha mãe Elci e minha namorada Aline.

Agradeço a minha orientadora, a professora Chang, por toda a disposição para prontamente me ajudar com qualquer dificuldade que tive ao longo dos estudos e da elaboração desse trabalho e por todos os seus ensinamentos.

Agradeço aos professores Ary, Hugo e Verónica por aceitarem compor a banca examinadora e pelas valiosas sugestões para a versão final deste trabalho. Em especial, ao professor Ary agradeço pelos cursos ministrados na graduação e no doutorado, além das valiosas conversas e plantões de dúvidas sobre integração estocástica. Também agradeço à professora Cátia, membro suplente da banca de avaliação, que foi a grande incentivadora da minha mudança de área entre graduação e mestrado, além de ter me orientado durante o mestrado.

Agradeço a todos que de alguma forma me ajudaram ao longo desses últimos anos, seja com discussões, conselhos ou, simplemente, companheirismo. Em especial à Alex, Marta, Léo, Nathália, Marcos, Filipe, Alancoc e Wenersamy. Aos companheiros de sala Pavel e Michel. Todos que me ajudaram nas prévias de qualificações e/ou me deram conselhos valiosos em outros quesitos do doutorado: Fábio, Geovane, Jailson, Márcio e Laena, Welington, Vinicius, Letícia, Lucas, Guillermo, Geraldo, Manuel, Marcos, Herlisvaldo, Gustavo, Irvin... Me desculpem aqueles que não mencionei, mas o tempo pra escrita dos agradecimentos foi curto demais para lembrar todos os nomes.

Também aos amigos do curso de bacharelado em Estatística Agda, Augusto, Érique, Felipe, Geisi, Guilherme, Márcia e Thiago, que muito me motivaram para este doutorado em Matemática. Em especial ao Luciano e Bruno por se disponibilizarem a rodar meus programas em servidores que possuíam acesso.

Agradeço aos professores e funcionários do departamento Matemática que colaboraram diretamente para minha formação. Professores Eduardo, Alberto, Leandro, Paulo e Daniele pelas discussões teóricas e valiosas dicas em aulas, seminários e conversas de corredores...

Agradeço ao Rafael e Giselly pelas revisões de textos e slides.

Por fim, agradeço à CAPES e ao CNPq pelo auxílio financeiro e ao SCI-HUB por tornar a ciência livre e acessível a todos.

Resumo

Apresentamos o estimador de máxima verossimilhança (MLE) para o parâmetro de *drift* de uma Equação de Langevin Generalizada (GLE) governada por um processo de Lévy observado continuamente no tempo. Em geral, o MLE não tem forma explícita e apresentamos condições suficientes para que o estimador seja consistente, assintoticamente normal e eficiente. Em particular, mostramos que o experimento estatístico associado à GLE satisfaz a propriedade de LAN (*locally asymptotic normal*). Propomos uma discretização do MLE utilizando filtro de grandes saltos (FMLE). Um segundo estimador discretizado é proposto usando as mesmas ideias do FMLE, mas introduzindo uma dependência do processo de Lévy simulado. Foram analisadas estimações de simulações do processo de Ornstein-Uhlenbeck generalizado do tipo exponencial flutuante com três parâmetros. Por fim, um caso particular do *drift* da GLE foi abordado, para o qual o MLE tem uma forma explícita e o FMLE herda as propriedades do estimador a tempo contínuo.

Palavras-Chave: Equação de Langevin Generalizada; Processo de Ornstein-Uhlenbeck Generalizado; Estimação do Drift; Propriedade de LAN; MLE Filtrado.

Abstract

We present the maximum likelihood estimator (MLE) for the drift parameter of the generalized Langevin equation (GLE) driven by a Lévy process observed continuously in time. Generally, the MLE has a non-explicit form and we present sufficient conditions for its consistency, asymptotic normality and efficiency. In particular, we show that the statistical experiment associated with the GLE satisfies the locally asymptotic normal (LAN) property. We propose a discretization of the MLE by filtering "big" jumps (FMLE). A second discretized estimator is proposed using the same ideas of the former, but introducing a path dependence of the simulated Lévy processes. Estimations from simulated paths were done for the 3-parameter generalized Ornstein-Uhlenbeck process of the fluctuating exponential type. Finally, a particular case of the GLE drift was considered, for which the MLE has an explicit form and the FMLE inherits the properties of the continuous time estimator.

Keywords: Generalized Langevin equation; Generalized Ornstein-Uhlenbeck processes; Drift estimation; LAN property; Filtered MLE.

Contents

List of Abbreviations and Symbols

The next list describes several symbols and abbreviations that will be later used within the body of the thesis

Abbreviations

GLE Generalized Langevin Equation

GOU Generalized Ornstein-Uhlenbeck

$GOU\text{-}FE$ Generalized Ornstein-Uhlenbeck of the Fluctuating Exponential type

LAN Locally Asymptotically Normal

MLE Maximum Likelihood Estimator

$SDDE$ Stochastic Delay Differential Equation

SDE Stochastic Differential Equation

$SIDE$ Stochastic Integro-Differential Equation

Estimators

$\hat{\theta}(t)$ Maximum Likelihood Estimator

$\hat{\theta}_T^{FMLE}$ Filtered Maximum Likelihood Estimator

$\hat{\theta}_T^{mFMLE}$ Modified Filtered Maximum Likelihood Estimator

Measures

ν Lévy measure

$\{\mu_{\theta,t}; (\theta, t) \in \Theta \times [0, \infty)\}$ Signed measures on $[0, t]$

$\{P_\theta; \theta \in \Theta\}$ Family of probability measures on $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\})$

P_θ^t The restriction of P_θ to the $\sigma-$algebra $\mathcal{F}_t$

$P_\theta \overset{loc}{\ll} P_{\theta_0}$ P_θ is locally absolutely continuous with respect to P_{θ_0}; if $P_\theta^t << P_{\theta_0}^t$ for all $t \in \mathbb{R}_+$

Processes

$b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds)$ Drift process of the GLE

$\frac{dP_\theta^t}{dP_0^t}(X_t)$ Likelihood function of $\mathbf{X}$; Radon-Nikodym derivative

$\mathbf{L} = \{L(t); t \geq 0\}$ Lévy process

$\mathbf{W} = \{W(t); t \geq 0\}$ Wiener process

$\mathbf{X} = \{X(t); t \geq 0\}$ GOU process; solution of a GLE

$l(\theta, X_t)$ Log-likelihood function

$X(t)$ Process value at time t

X^c Continuous martingale part of $\mathbf{X}$

$X_t = \{X(s); s \leq t\}$ Past information of $\mathbf{X}$ until the time t

Other Symbols

(b, σ^2, ν) Characteristics triplet of a Lévy process

$(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\})$ Filtered probability space

$\mathbb{R}_+$ Positive real numbers $(0, \infty)$

$\mathcal{F}$ $\sigma-$algebra on Ω

Law $(V(t) \mid \mathbb{P})$ The distribution of the random vector $V(t)$ under the measure $\mathbb{P}$

Law $(V(t) \mid \mathbb{P}) \to V$ The convergence in distribution under the measure $\mathbb{P}$

$\nabla f : \mathbb{R}^N \to \mathbb{R}^N$ Gradient vector of a function f

$\partial^2 f$ Hessian matrix of a function f

$\partial_i f$ Partial derivative of a function f

$\rho'_\theta(t)$ $\frac{d}{dt}\rho_\theta(t)$

$\rho_\theta(t)$ Resolving function of the GOU process

$\mathbf{A}^\top$ Transpose of a matrix $\mathbf{A}$

$\mathbf{v}^\top$ Transpose of a vector $\mathbf{v}$

Θ Parameter space; subspace of $\mathbb{R}^N$

$\{\mathcal{F}_t; t \geq 0\}$ Filtration of sub-$\sigma-$algebras on Ω

$a \wedge b$ Minimum of a and b

$D[t_0, \infty)$ Càdlàg functions on the interval $[t_0, \infty)$

E_θ Expectation with respect the measure P_θ

$f * g(t)$ Convolution of the functions f and g

$H * \mu$ Integral of H with respect to the measure μ

$H \cdot A$ Integral of H with respect to the process A. Defined in (A.1)

I_N $N \times N$-identity matrix

$L^p\left(dP_\theta \times dt\right)$ Functions $p-$integrables with respect to the product measure

$S_\alpha\left(\sigma, \beta, \mu\right)$ Stable distribution with index α, scale parameter σ, skewness parameter β and shift parameter μ

Introduction

Applications of *Stochastic Differential Equations* (SDEs) have been developed in many fields such as physics, economics and chemistry since early 20th century. More recently, great interest has been arisen in models with some memory effects, where *Stochastic Integro-Differential Equations* (SIDEs) can play an important role. An example of SIDE is the *Generalized Langevin Equation* (GLE), proposed in the 1960s by Kubo [29] and Mori [51]. In such SIDE, the drift component takes into account the evolution of the process up to a considered time t. Formally, the GLE can be written as

$$\begin{cases} dX(t) &= -\int_0^t X(s)\gamma(t-s)ds\,dt + dL(t), \quad t > 0 \\ X(0) &= X_0, \end{cases} \tag{1}$$

where $\mathbf{L} = \{L(t); t \geq 0\}$ is a stochastic noise, X_0 is a random variable independent of $\mathbf{L}$, $\gamma(\cdot)$ is a deterministic function called *Memory Function* and the drift is $-\int_0^t X(s)\gamma(t-s)ds$. It is worthwhile to mention that if the memory function is a Dirac delta concentrated at the constant $\theta > 0$ and the noise is the Wiener process $\mathbf{W} = \{W(t); t \geq 0\}$, then the GLE becomes the widely known *Classical Langevin Equation*

$$\begin{cases} dX(t) &= -\theta X(t)dt + dW(t), \quad t > 0, \\ X(0) &= X_0, \end{cases} \tag{2}$$

whose solution is the so called *Classical Ornstein-Uhlenbeck* (OU) process and is given by

$$X(t) = X_0 e^{-\theta t} + \int_0^t e^{-\theta(t-s)} dW(s), \quad t \geq 0.$$

Results on existence and uniqueness of solution for certain classes of the GLE were obtained by Kannan [24], Kannman and Bharucha-Reid [25], Medino et al. [49] and Santos [59]. Along the last two decades, further theoretical matters and applications of the GLE have been performed by many authors such as Di Terlizzi et al. [9], McKinley and Nguyen [48], Ottobre and Pavliotis [52], Pravliotis et al. [53], Zhu and Venturi [66] and Zwanzing [67].

Concerning statistical inference issues, Küchler and Sϕrensen [33] investigate the maximum likelihood estimation problem for the drift parameter of the classical Langevin equation only. On the other hand, substituting the Wiener process in the noise term by a more general Lévy one in the classical Langevin equation, non-Gaussian Ornstein-Uhlenbeck processes can be obtained as solution. This is the case in the famous work of Barndorff-Nielsen and Shephard [3], who has been considered as building block for important applications in finance. In such setting, Mai [42, 43] deals with maximum likelihood estimation matters for classical OU processes driven by Lévy noises.

Issues concerining statistical estimation for the drift term of the GLE are important to better fit the mathematical model to the potential real time series aimed to be modeled. They enable us to infer predictions on future matters of the modeled time series based on accumulated information.

Statistical inference issues concerning the GLE also deserve attention since the relevance that such equation can play in modelling phenomena with memory effects. Stein, Lopes and Medino [62] present an alternative approach to the GLE which generalizes the previously used one. They consider a family of finite signed measures $\{\mu_t; t \geq 0\}$ instead of a memory function $\gamma(\cdot)$. With this procedure, the GLE becomes

$$\begin{cases} dX(t) &= -\int_0^t X(s)\mu_t(ds)dt + dL(t), \quad t > 0 \\ X(0) &= X_0. \end{cases}$$

Then, theoretical properties such as recurrence and autocorrelation structure of the solution processes from classes of GLE were investigated, as well as simulation results were presented. For instance, a case introduced by the authors, named Cosine Process, present non-stationary behaviour, with fluctuating paths and oscillating decay for the autocorrelation function.

Inspired by ideas in Stein, Lopes and Medino [62], Alcântara [1] introduced a GLE based model for domestic dollar interest rates time series, known as *Cupom Cambial*. A drawback in modelling Cupom Cambial by classic OU processes is that their embedded time series exhibit characteristics of order 1 autoregressive models (AR(1)) and this does not fits to real data evidences. Thus, the author propose a class of GLE based model which better adapts to Cupom Cambial time series. Briefly, such class of model corresponds to a mixture of classical OU process and cosine process.

In this thesis we investigate properties from the *Maximum Likelihood Estimator* (MLE) for the drift parameter of classes of GLE. That is, we consider a pre-specified vector parameter θ on which the memory function $\gamma(\cdot)$ depends and a Lévy process $\mathbf{L}$ as the driving noise. We present sufficient conditions for the consistency of the MLE (Law of Larger Numbers-LLN) as well as the asymptotic normality (Central Limit Theorem-CLT).

Having established results on consistency and on asymptotic normality for the MLE, a natural improvement is to search for information on the efficiency property of the estimator. In this direction, we prove that the MLE is efficient in the sense of the Hájek-Le Cam Convolution Theorem, extending results obtained by Mai [42, 43] for Lévy-driven OU processes. Precisely, we show that the statistical experiment associated with the GLE satisfies the *Locally Asymptotic Normal* (LAN) property (see Theorem 2.3.4). Such notion of efficiency is directly related to statistical experiments satisfying the LAN property and avoids super-efficiency issues (cf. Ibragimov and Has'minskii [22]).

The LAN property is a useful instrument to study efficiency of statistical estimators. It was introduced by Le Cam [35] and has been widely studied ever since. We refer the reader to Ibragimov and Has'minskii [22], Le Cam and Yang [36] and Vaart [65] for further details on this subject. Recent studies on the efficiency of estimators considering the LAN property have been performed in the specific case of SDEs, see for example Benke and Pap [5, 6], Gloter, Loukianova and Mai [13], Gushchin and Küchler [14, 16], Kohatsu-Higa, Nualart and Tran [28], Liu, Nualart and Tindel [39], Mai [42, 43] and Tran [63, 64].

As we have mentioned, Kannan [24], Kannman and Bharucha-Reid [25], Medino et al. [49] and Santos [59] explored issues on existence and uniqueness of solutions for certain classes of GLE. Under mild conditions, a representation form for such solutions was first presented by Kannan [24] as a stochastic process $\{X(t); t \geq 0\}$ given by

$$X(t) = X_0\rho(t) + \int_0^t \rho(t-s)dL(s),$$

where $\rho(\cdot)$ is a deterministic function satisfying the following Volterra integro-differential equation

$$\begin{cases} \rho'(t) &= -\int_0^t \rho(s)\gamma(t-s)ds, \\ \rho(0) &= 1\cdot \end{cases} \tag{3}$$

Such representation form is called *Generalized Ornstein-Uhlenbeck* process (GOU) and we recall that the classical OU process is the particular case corresponding to the exponential $\rho(t) = e^{-\theta t}$.

In this work, we present a detailed study for the Generalized Ornstein-Uhlenbeck of Fluctuating Exponential type (GOU-FE) process. This class of solution was proposed by Alcântara [1] to model Cupom Cambial time series. In the Alcântara's work, a representation form is considered with function ρ given by

$$\rho(t) = \theta_2 e^{-\theta_1 t} + (1-\theta_2)\cos(\theta_1 t),$$

where $(\theta_1, \theta_2) \in \mathbb{R}^2$ is a two dimensional parameter to determine. As we also have pointed out, Alcântara's model is a mixture of the classical OU process and the cosine process studied in Stein, Lopes and Medino [62].

Our strategy is to allow the behavior of the exponential and cosine components depends on different parameters. That is, we consider a GOU-FE process with

$$\rho(t) = \theta_3 e^{-\theta_1 t} + (1-\theta_3)\cos(\theta_2 t) \tag{4}$$

for $(\theta_1, \theta_2, \theta_3) \in \Theta \subset \mathbb{R}^2 \times [0, 1]$. Similarly to Alcântara, we prove that the proposed process is a GLE solution, that is, we found explicitly $\gamma(\cdot)$ satisfying (3). We also show that the discretized process behaves as an order 3 autoregressive process. Through simulation studies, we evaluate the MLE's performance in estimating the parameters in GOU-FE process as well as in our modification.

To carry out a simulation study of the MLE performance, a discretized form of the likelihood is required. The main problem in the discretization of the MLE is getting a good approximation for the increments of the continuous martingale part X^c which is not observable. We discretized the MLE through filtering "big" jumps of the process and we will call it FMLE (filtered MLE). The motivation for this approach was Mancini [44, 45] –who studied the order of the X^c increments and developed a technique for identifying the times when jumps larger than a defined threshold occur– and Mai [43] and Gloter, Loukianova and Mai [13] –who used a thresholding technique to approximate the continuous part. A second discretized estimator was proposed using the same ideas of the former, but introducing a path dependence of the simulated Lévy processes. The discretized estimators were compared using GOU-FE process simulations with $\rho(\cdot)$ defined by (4).

To study the asymptotic behavior of the FMLE, we restrict our investigation to the GLE given by

$$\begin{cases} dX(t) &= -\sum_{j=1}^{N} \theta_j \int_0^t X(s)\gamma_j(t-s)dsdt + dL(t), \quad t > 0 \\ X(0) &= X_0, \end{cases} \tag{5}$$

where $\theta^\top = (\theta_1, \cdots, \theta_N) \in \Theta \subset \mathbb{R}^N$. With this restriction, under suitable regularity conditions, we find an explicit form from the MLE for θ and we prove that the MLE inherits the same asymptotic behavior of the MLE in the general case. We also show that the FMLE has the same asymptotic behaviour of the theoretical estimator MLE.

This work is organized in four Chapters and one Appendix. In Chapter 1, we present basic notations, concepts and properties that will be used in the development of the subsequent chapters. In particular, we present the formalization of LAN property and a summary of recent statistical inference studies on stochastic processes, especially those that are of our particular interest.

In Chapter 2, we explore the connection between the probability measures $\{P_\theta; \theta \in \Theta\}$ (induced by $\mathbf{X}$) and the Radon-Nikodym derivative of P_θ with respect to P_0 which allows us to write the log-likelihood function as

$$l(\theta, X_t) := \log \frac{dP_\theta^t}{dP_0^t}(X_t) = \frac{1}{\sigma^2}\int_0^t b(\theta, X_s)dX^c(s) - \frac{1}{2\sigma^2}\int_0^t b^2(\theta, X_s)ds, \tag{6}$$

where $X_t = \{X(s); 0 \leq s \leq t\}$, $b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds)$ and X^c denotes the continuous martingale part of X. Though, in general, we do not have an explicit form for the MLE $\{\hat{\theta}(t); t \geq 0\}$, we can bypass this difficulty by using versions of CLT for continuous-time multivariate martingales. Our Theorem 2.3.4 shows that, under ergodicity assumptions and regularity conditions, $\{\hat{\theta}(t); t \geq 0\}$ is strongly consistent and asymptotically normal. Furthermore, we show that $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN property, which in turns assures that $\{\hat{\theta}(t); t \geq 0\}$ is asymptotically efficient in the sense of Hájek-Le Cam convolution theorem, that is

$$\text{Law}\left(\varphi(t)^{-1}\left(\hat{\theta}(t) - \theta\right)\Big| P_\theta\right) \to I(\theta)^{-1/2}N\left(0, I_N\right) \ \text{ as } \ t \to \infty,$$

where $\varphi(t)$ is an appropriate normalizing function, N denotes the normal random vector with covariation matrix I_N, $I(\theta)$ is the asymptotic Fisher information matrix, and Law $(V(t) \mid P_\theta) \to V$ denotes the convergence in distribution under P_θ. To overcome difficulties that arise from the discretization X^c we make use of FMLE that filters the "big" jumps of the process $\mathbf{X}$, that is

$$\hat{\theta}_T^{FMLE} = \arg\max_{\theta \in \Theta_0}\left\{\frac{1}{\sigma^2}\sum_{i=1}^n b(\theta, X_{t_i})\Delta_i X \mathbf{1}_{[|\Delta_i X| \leq v_n^i]} - \frac{1}{2\sigma^2}\sum_{i=1}^n b^2(\theta, X_{t_i})\Delta_i\right\}.$$

Simulation results of FMLE for some GOU-FE models are exhibited.

Chapter 3 is dedicated to study the GOU process of Fluctuating Exponential type with function $\rho_\theta(\cdot)$ given by (4). This process will allow us to model phenomena that present oscillations or seasonality or autoregressive behavior. Our simulation results for the process

$$X(t) = X_0(0)\left((1-\theta_2)e^{-\theta_1 t} + \theta_2\cos(\theta_2 t)\right) + \int_0^t \left((1-\theta_2)e^{-\theta_1(t-s)} + \theta_2\cos(\theta_2(t-s))\right)dL(s)$$

shows that despite the good results obtained when the perturbation of the OU process is small, i.e., $\theta_2 \downarrow 0$, issues such as computational time and improved estimation in some regions of parametric space require a little more care when θ_2 is not small. This led us to propose a more general "memory" function by introducing the parameter θ_3 as in (4). Worth pointing out that this process will allow us to model phenomena that present oscillations or seasonality or higher autoregressive behavior. In fact, our Proposition 3.2.2 shows that a discretization of this new process has an order 3 autoregressive form, which is a consequence of the OU and cosine recurrences. Moreover, based on the recurrent structure of the model and making use of the simulated noise and the real expected drift, a new modified FMLE is proposed (Section 3.3),

$$\hat{\theta}_T^{mFMLE} = \arg\min_{\theta\in\Theta_0} |l(\theta, X_{t_n}) - l(\theta_0, X_{t_n})|,$$

where θ_0 is the true value of the unknown parameter and $\{t_1, \cdots, t_n\}$ is a convenient partition of the time interval. Simulation results for mFMLE and its performance is compared to that of FMLE in Section 3.3. Our Theorem 3.2.1 shows that the corresponding process is indeed a solution of the GLE with the associated family of signed measures $\{\mu_{\theta,t}; \theta \in \Theta,\ t \geq 0\}$ satisfying the Volterra integro-differential equations (3). Also, the Theorem 3.2.6 shows that if the "memory" function $\rho_\theta(\cdot)$ satisfies a recurrence relation, then the corresponding GOU process will have autoregressive representation of general order m. Thus the same discretization technique could be applied. Extension for the case when $\mathbf{L}$ is a symmetric $\alpha-$stable Lévy process with $1 < \alpha \leq 2$ is considered in Proposition 3.2.7.

In Chapter 4, we will consider a particular class of GLE for which the drift function satisfies

$$b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds) \text{ and } \mu_{\theta,t} = \sum_{j=1}^{N} \theta_j \mu_t^{(j)},$$

where the signed measures $\mu_t^{(j)}$ are finite at intervals $[0, t]$. These restrictions appear naturally when we analyze the OU, cosine and stochastic delay (SDDE) processes. For this type of processes the parameter θ can be linearly separated and depends exclusively on the process history X_t. Moreover, for this class of GLE the Radon-Nikodym density of P_θ with respect to P_0 takes up a much simpler form and we can derive an explicit expression for the MLE,

$$\hat{\theta}(t) := \arg\max_{\theta\in\Theta_0} \frac{dP_\theta^t}{dP_0^t}(X_t) = S(t)^{-1}A(t),$$

where the matrices $S(t)$ and $A(t)$ are properly defined functions of the drift function $b(\theta, X_t)$ and the continuous martingale part X^c. Based on this and under much milder assumptions than those of Theorem 2.3.4, our Theorem 4.2.4 establishes for $\hat{\theta}(t)$ the strong consistency, asymptotic normality and asymptotic efficiency in the sense of Hájek-Le Cam Convolution Theorem. Its proof is less technical and basically requires a version of the CLT for continuous time martingales. In Section 4.3, we study the asymptotic behavior of the corresponding discretized and filtered approximation $\hat{\theta}_n^{FMLE} := S_n^{-1}A_n$ where S_n and A_n, are approximations of $S(t)$ and $A(t)$ respectively. As shown in Theorem 4.3.2, with a good convergence rate for some stochastic integrals approximations by Riemann sums, the FMLE will have the same asymptotic behavior as that of MLE.

For the convenience of the reader and aiming a work as self-contained as possible, in Appendix A we recall some concepts from the theory of stochastic processes and statistical inference we have used through the Thesis. We also present a survey of results from related publications which are fundamental for the development of this work.

Chapter 1

Preliminaries

1.1 Introduction

In this chapter, we present preliminary concepts that are used throughout this thesis. We start with Section 1.2 setting the basic notations, defining the *Locally Asymptotically Normal* (LAN) property for a statistical experiment associated with a continuous time stochastic process. We also present examples and establish the connection between LAN property and the notion of efficiency for estimators, in the sense of Hájek-Le Cam convolution theorem.

Section 1.3 is dedicated to present some classes of Lévy-driven stochastic differential equations and its recent statistical inference results. These processes are particular cases of the *generalized Langevin equation* (GLE) which is presented in Definition 1.3.7 using the approach of Stein, Lopes and Medino [62], that is, changing the memory kernel of the Mori [51] and Kubo [29] definition by a family of signed measures. The sense of Kannan's solution for the GLE named *generalized Ornstein-Uhlenbeck process* (see [1, 24, 49, 59, 62]) is also presented. Although it is not a particular case of GLE, the last class of processes presented (which was studied by Gloter, Loukianova and Mai [13]) gives us an idea of techniques to be used to obtain the behavior of the drift estimator in Chapter 2. In Section 1.4 we state the problems that will be studied in Chapters 2, 3 and 4.

Finally, we emphasize that the reader who is familiar with statistical inference for Stochastic Differential Equations can skip this chapter and go straight to Chapter 2, referring to this chapter only for occasional citations.

1.2 Notations and Basic Definitions

We start this section by setting some notations. Let us denote the distribution of the random vector $V(t)$ under the measure $\mathbb{P}$ by $\text{Law}\,(V(t) \mid \mathbb{P})$ and

$$\text{Law}\,(V(t) \mid \mathbb{P}) \to V, \quad \text{as} \quad t \to \infty,$$

denotes the convergence in distribution.

Let $\mathbf{A} = (a_{jk})_{N\times N} \in \mathbb{R}^{N\times N}$ be a positive semi-definite $N \times N$-matrix. Denote $\mathbf{A}^{1/2}$ and $\det(\mathbf{A})$ its positive semi-definite square root and its determinant, respectively. Let $\mathbf{v}^\top = (v_1, \cdots, v_n) \in \mathbb{R}^N$ be a vector. We denote

$$(\text{diag}\,\mathbf{A})^\top = (a_{11}, a_{22}, \cdots, a_{NN})$$

the diagonal vector of $\mathbf{A}$ and diag $\mathbf{v}$ the diagonal $N \times N$-matrix with $\mathbf{v}$ as diagonal.

For $f : \Theta \to \mathbb{R}$ twice continuous differentiable, we denote its *gradient column vector* and the *Hessian matrix*, respectively, by $\nabla f : \Theta \to \mathbb{R}^N$ and $\partial^2 f = \left(\partial^2_{\theta_i \theta_j} f\right)_{N\times N}$.

Consider P_θ and P_{θ_0} two probability measures on the filtered measure space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\})$. We say that P_θ is locally absolutely continuous with respect to P_{θ_0} and write $P_\theta \overset{loc}{\ll} P_{\theta_0}$ if $P^t_\theta << P^t_{\theta_0}$ for all $t \in \mathbb{R}_+$, where $P^t_\theta := P_\theta|_{\mathcal{F}_t}$ denotes the restriction of P_θ to $\mathcal{F}_t$.

1.2.1 Locally Asymptotically Normal Property

An important condition to study the asymptotic efficiency of estimators is the *Locally Asymptotically Normal* (LAN) property. This theory was introduced by Le Cam [35] and has been widely studied ever since. Several authors have shown that, under the ergodicity condition, the statistical experiment $\{P_\theta; \theta \in \Theta\}$ associated with a stochastic process $\mathbf{X} = \{X(t); t \geq 0\}$ satisfies the LAN property (see for example [13, 32, 42, 63, 64]).

In this section we summarise the importance of the LAN property and the notion of asymptotic efficiency used in the main results of Chapters 2 and 4. For a more detailed study on this subject we suggest [22, 36, 65]. We start with a basic motivation.

Let $X_1, \cdots, X_n$ be a sequence of independent and identically distributed random variables on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\Theta \subset \mathbb{R}$ be a parameter space.

Consider a parametric statistical experiment $(\mathcal{X}^n, \mathbb{B}(\mathcal{X}^n), \{P^n_\theta \in \Theta\})$ such that

$$\mathbf{X} = (X_1, \cdots, X_n) : \Omega \times \Theta \to \mathcal{X}^n.$$

Here P^n_θ is the probability measure on $(\mathcal{X}^n, \mathbb{B}(\mathcal{X}^n))$ induced by $\mathbf{X}$ under θ.

We are interested in finding an estimator $T_n : \mathcal{X}^n \to \Theta$ with a "good" asymptotic behavior. Denote the *Maximum Likelihood Estimator* (MLE) by

$$\hat{\theta}_n = \hat{\theta}_n(X_1, \cdots, X_n) := \arg\max_{\theta\in\Theta} \frac{dP^n_\theta}{d\nu^n}(X_1, \cdots, X_n).$$

Here the likelihood is the Radon-Nikodym derivative of the measure P^n_θ with respect to the Lebesgue measure ν^n (or the counting measure in discrete experiments). In general, under suitable regularity conditions, it follows from the LLN (law of large numbers) and CLT (central limit theorem) that the MLE is consistent and asymptotically normal.

A natural question that we would like to answer is: If $T_n : \mathcal{X}^n \to \Theta$ is another sequence of estimators for θ with similar asymptotic behavior, which one should we use?

A famous result in statistical inference that helps us to answer this question is the following. Its proof can be found, for example, in Ibragimov and Has'minskii [22, Theorem I.7.3].

Theorem 1.2.1 (The Cramér-Rao Inequality)**.** *Let $\{P_\theta; \theta \in \Theta\}$ be a regular parametric statistic model and the statistic $T : \mathcal{X} \to \mathbb{R}^k$ is such that $E_\theta\left(|T-\theta|^2\right)$ is bounded in a neighborhood of the point $\theta \in \Theta$. Assume that the Fisher Information matrix $I(\theta)$ is invertible for all $\theta \in \Theta$. Then the bias*

$$b(\theta) = E_\theta T - \theta$$

is continuously differentiable in this neighborhood of θ and the following inequality is satisfied

$$E_\theta (T-\theta)(T-\theta)^\top \geq \left(J + \frac{\partial b(\theta)}{\partial \theta}\right) I(\theta)^{-1} \left(J + \frac{\partial b(\theta)}{\partial \theta}\right)^\top + b(\theta)b(\theta)^\top,$$

where J denotes the unit matrix.

In particular, if T is an **unbiased estimator** *of θ, then*

$$E_\theta (T-\theta)(T-\theta)^\top \geq I(\theta)^{-1}. \tag{1.1}$$

This result gives us an idea to study estimators with minimal variance. The following definition makes this motion more precise.

Definition 1.2.2. *1. We say that an estimator $T : \mathcal{X} \to \Theta$ is* **efficient** *if the equality in (1.1) is satisfied.*

2. An estimator $\hat{\theta}_n : \mathcal{X}^n \to \Theta$ is called **asymptotically efficient in the Cramér-Rao sense** *if it is asymptotically normal, and its covariance matrix achieves asymptotically the Cramér-Rao lower bound.*

Remark 1.2.3. *As described by Ibragimov and Has'minski [22], the term* **asymptotically efficient estimator** *was introduced by R. Fisher to describe asymptotically normal estimators with minimal asymptotic variance.*

The Fisher's idea was to show that the MLE $\hat{\theta}_n$, under natural regularity conditions, satisfies that $\sqrt{n}\left(\hat{\theta}_n - \theta\right)$ is asymptotically normal with parameters $(0, I(\theta)^{-1})$ and if T_n is another sequence of asymptotically normal estimators, then

$$\lim_{n\to\infty} E_\theta \left[n\left(T_n - \theta\right)^2 \right] \geq I(\theta)^{-1}, \quad \theta \in \Theta.$$

Example 1.2.4 (The well known Hodges' counterexample)**.** *Let $X_1, \cdots, X_n$ be a random sample of $N(\theta, 1)$. Consider $\hat{\theta}_n = \bar{X}_n$ the MLE of θ and set*

$$T_n = \begin{cases} \bar{X}_n, & \left|\bar{X}_n\right| > n^{-1/4}, \\ 0, & \left|\bar{X}_n\right| \leq n^{-1/4}. \end{cases}$$

Then T_n is a **super-efficient** *estimator for $\theta = 0$ (in the sense of Ibragimov and Has'minskii [22, Equation (I.9.1)]).*

Le Cam [35] in the 60s introduced a concept that combined with Hájek [18, 19] in the 70s allows us to redefine the class of efficient estimators. Consider the Radon-Nikodym derivative of the measure P_θ with respect to the Lebesgue measure ν (or the counting measure in discrete experiments)

$$f(x;\theta) = \frac{dP_\theta}{d\nu}(x), \quad x \in \mathcal{X}\cdot$$

Let

$$Z_{n,\theta}(h) = \prod_{j=1}^{n} \frac{f(X_j; \theta + h/\sqrt{n})}{f(X_j;\theta)}$$

denotes the normalized likelihood ratio.

Theorem 1.2.5 (L. Le Cam)**.** *Under suitable regularity conditions, for all $h \in \mathbb{R}^N$,*

$$Z_{n,\theta}(h) = \exp\left\{ \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \left(\frac{\partial}{\partial\theta} \log f\left(X_j;\theta\right), h \right) - \frac{1}{2}\left(I(\theta)h, h\right) + o_{P_\theta^n}(1) \right\}$$

where

$$\text{Law}\left(\frac{1}{\sqrt{n}}\sum_{j=1}^{n}\frac{\partial}{\partial\theta}\log f\left(X_j;\theta\right)\middle| P_\theta^n\right) = N(0, I(\theta)).$$

Here $(\cdot,\cdot)$ *denotes the Euclidean inner product.*

Proof. See Ibragimov and Has'minskii [22, Theorem II.1.1, p.114-117].

After this result, it becomes natural to study statistical experiments in which the likelihood ratio satisfies (locally) a quadratic expression like in Theorem 1.2.5, but with less restrictive dependence between the random variables associated.

Below, we present the notion of asymptotic efficiency in the sense of Hájek-Le Cam Convolution Theorem in the context of continuous times stochastic processes (see Ibragimov and Has'minskii [22, Theorem II.9.1], Le Cam and Yang [36, Theorem 6.3] or Vaart [65, Theorem 8.8]). For this, we first present the definition of LAN statistical experiment and some examples of stochastic process satisfying this property and then we state the asymptotic efficiency.

Definition 1.2.6. *We say that a parametric statistical experiment* $\{P_\theta;\theta\in\Theta\}$ *on* $\big(\Omega,\mathcal{F},\{\mathcal{F}_t;t\geq 0\}\big)$ *satisfies the* ***Locally Asymptotic Normal (LAN) property*** *at* $\theta\in \text{int}\,\Theta$ *if there exists an invertible* $N\times N$*-matrix* $\varphi(t)$ *and a positive definite* $N\times N$*-matrix* $I(\theta)$ *such that, for any* $h\in\mathbb{R}^N$*, the following limit holds true*

$$\text{Law}\left(\log\frac{dP^t_{\theta+\varphi(t)h}}{dP^t_\theta}\middle| P_\theta\right)\to h^\top N\left(0,I(\theta)\right)-\frac{1}{2}h^\top I(\theta)h,\ \ \text{as}\ \ t\to\infty, \tag{1.2}$$

where $N\left(0,I(\theta)\right)$ *is a* N*-dimensional Gaussian random vector with covariance matrix* $I(\theta)$*.*

At the reader convenience the remaining of this subsection may be skipped for latter reading. Essentially, we present a list of examples of processes satisfying Definition 1.2.6 and define the notion of efficiency based on the LAN property.

Remark 1.2.7. *Note that, in Definition 1.2.6, the vector* $\theta+\varphi(t)h$ *does not have to belong to* Θ *for all* $t>0$*. However, since* $\theta\in\text{int}\,\Theta$*, it will belong to* Θ *when* t *is large enough, for each* $h\in\mathbb{R}^N$*.*

An equivalent form to define the LAN property is stating when the statistical experiment $\{P_\theta;\theta\in\Theta\}$ is locally quadratic and mixed normal. This definition is particularly interesting because when $\{P_\theta;\theta\in\Theta\}$ does not satisfies the LAN property, we can still study the local properties of the Radon-Nikodym derivatives (see, for example, Gushchin and Küchler [16, Tables 1 and 2]).

Consider a statistical experiment $\{P_\theta;\theta\in\Theta\}$ on the filtered space $\big(\Omega,\mathcal{F},\{\mathcal{F}_t;t\geq 0\}\big)$.

Definition 1.2.8. *We say that* $\{P_\theta;\theta\in\Theta\}$ *is*

1. ***Locally Asymptotic Quadratic (LAQ)*** *at* $\theta\in\text{int}\,\Theta$ *if there exists a* N*-dimensional random process* $\{Z_\theta(t);t\geq 0\}$ *and a family* $\{I_\theta(t);t\geq 0\}$ *of symmetric positive semi-definite random* $N\times N$*-matrices such that*

$$\log\frac{dP^t_{\theta+\varphi(t)h}}{dP^t_\theta}-h^\top I_\theta(t)^{1/2}Z_\theta(t)+\frac{1}{2}h^\top I_\theta(t)h=o_{P_\theta}(1),$$

for every $h\in\mathbb{R}^N$*, where* $\{\varphi(t);t\geq 0\}$ *is a family of positive definite matrices satisfying* $\varphi(t)\to 0$ *as* $t\to\infty$*;*

2. *Locally Asymptotically Mixed Normal (LAMN) at $\theta \in int\,\Theta$ if it is LAQ and*

$$Law\big(Z_\theta(t), I_\theta(t) \mid P_\theta\big) \to \big(Z_\theta, I_\theta\big) \text{ as } t \to \infty.$$

Here I_θ is a P_θ-almost sure positive definite $N \times N$-matrix such that, conditionally on I_θ, Z_θ is a N-dimensional random vector standard normal distributed;

3. *Locally Asymptotic Normal (LAN) at $\theta \in int\,\Theta$ if it is LAMN and I_θ is deterministic.*

Examples of continuous time stochastic processes satisfying the LAN property can be found in Benke and Pap [5, Proposition 4.1], Benke and Pap [6, Theorem 3.1], Gloter, Loukianova and Mai [13, Theorem 5.3], Gushchin and Küchler [16, Proposition 2.1], Ibragimov and Has'minskii [22, Theorem II.7.1], Liu, Nualart and Tindel [39, Theorem 1.7], Mai [42, Theorem 4.2.7] and Mai [63, Theorem 2.4.2]. We present some of them bellow.

Example 1.2.9 (Theorem II.7.1 in Ibragimov and Has'minskii [22])**.** *Consider a process of observations defined by the stochastic equation*

$$dX(t) = S(t,\theta)dt + dW(t), \quad 0 \le t \le T,$$

where $W(t)$ is a standard Wiener process, $\int_0^T S^2(t,\theta)dt < \infty$, $\theta \in \Theta \subset \mathbb{R}^N$. The measures $\{P_\theta^T; \theta \in \Theta\}$ are absolutely continuous for different θ and satisfies

$$\frac{dP_\theta^T}{dP_{\theta_0}^T} = \exp\left\{\int_0^T \left(S(t,\theta) - S(t,\theta_0)\right) dX(t) - \frac{1}{2}\int_0^T \left(S(t,\theta) - S(t,\theta_0)\right)^2 dt\right\}.$$

Then $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN condition at all $\theta \in int\,\Theta$.

Example 1.2.10 (Mai [42, 43])**.** *Consider a Ornstein-Uhlenbeck process*

$$X(t) = e^{-\theta t}X(0) + \int_0^t e^{-\theta(t-s)}dL(s), \quad t \in \mathbb{R}_+,$$

where $L(t)$ is a Lévy process with Lévy-Khintchine triplet (b, σ^2, μ) (cf. Appendix A.3) and $\theta \in \Theta = \mathbb{R}_+$.

The Ornstein-Uhlenbeck process exhibits a modification with càdlàg paths and hence it induces a measure P_θ on the space $D[0,\infty)$ of càdlàg functions.

The likelihood function is given by (cf. [42, Proposition 3.2.4])

$$\frac{dP_\theta^t}{dP_0^t}(X_t) = \exp\left\{-\frac{\theta}{\sigma^2}\int_0^t X(s)dX^c(s) - \frac{\theta^2}{2\sigma^2}\int_0^t X^2(s)ds\right\}$$

Then, under the existence of solution and ergodicity conditions, $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN condition at all $\theta \in int\,\Theta$.

Example 1.2.11 (Gloter, Loukianova and Mai [13])**.** *Consider the process given by*

$$\begin{aligned} X(t) &= X(0) + \int_0^t b(\theta, X(s))ds + \int_0^t \sigma(X(s))dW(s) \\ &\quad + \int_0^t \int_{\mathbb{R}} \gamma(X(s^-))zN(ds,dz), \quad t \in \mathbb{R}_+, \end{aligned}$$

where $W(t)$ is a one-dimensional Brownian motion and $N(.,.)$ is the Poisson random measure (cf. Applebaum [2, Section 2.3]) on $\mathbb{R}_+ \times \mathbb{R}$ associated with the jumps of the Lévy process.

The log-likelihood function is defined as

$$l(\theta, X_t) = \exp\left\{\int_0^t \frac{b(\theta, X(s))}{\sigma^2(X(s))} dX^c(s) - \frac{1}{2}\int_0^t \frac{b^2(\theta, X(s))}{\sigma^2(X(s))} ds\right\}.$$

Then, under the existence of solution and ergodicity conditions, $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN condition at all $\theta \in$ int $\Theta \subset \mathbb{R}^d$.

Example 1.2.12 (Tran [63]). *As the previous examples, similar results are obtained for the processes $\mathbf{X} = \{X(t); t \geq 0\}$ that are solution (via Malliavin Calculus) for the following stochastic integral equation*

$$X_\theta(t) = \int_0^t b(\theta, X(t))dt + \int_0^t \sigma(X(t))dW(t) + \int_{\mathbb{R}_0} \gamma(X(t^-), z)\big(N(dt, dz) - \nu(dz)dt\big).$$

To redefine the notion of efficiency, as mentioned above, studies on Hájek [18, 19] are very important. We started presenting a version of the Hájek-Le Cam Convolution Theorem which was stated in Küchler and Sϕrensen [33, Theorem 8.5.2]. First, recall an experiment to satisfy the LAMN property in Definition 1.2.8.

Theorem 1.2.13 (The Hájek-Le Cam Convolution Theorem). *Suppose that the filtered statistical space $\big(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \{P_\theta; \theta \in \Theta\}\big)$ satisfies the LAMN property at $\theta \in$ int Θ and let $\left\{\hat{\theta}(t); t \geq 0\right\}$ be a family of estimators for θ that are $\mathcal{F}_t$-measurable and for each $h \in \mathbb{R}^N$*

$$Law\left(\varphi^{-1}(t)\left(\hat{\theta}(t) - \theta - \varphi(t)h\right)\middle| P_{\theta+\varphi(t)h}\right) \to V(\theta)$$

as $t \to \infty$, for some N-dimensional random vector $V(\theta)$. Then, there exists a stochastic kernel $K_{I(\theta)}$ such that

$$Law\,(V(\theta) \mid I(\theta)) = K_{I_\theta} * N\left(0, I(\theta)^{-1}\right),$$

where $*$ denotes the convolution function.

Proof. See Ibragimov and Has'minskii [22, Theorem II.9.1], Le Cam and Yang [36, Theorem 6.3], Vaart [65, Theorem 8.8].

As summarized in Tran [63], the Theorem 1.2.13 guarantees that the random vector $V(\theta)$ can be decomposed as a sum of two independent random vectors, i.e.,

$$V(\theta) \overset{d}{=} I(\theta)^{-1/2}N\left(0, I_N\right) + R,$$

where R is a random vector independent of $N\left(0, I_N\right)$. For $h = 0$, the Convolution Theorem implies that

$$\text{Law}\left(\varphi^{-1}(t)\left(\hat{\theta}(t) - \theta\right)\middle| P_\theta\right) \to I(\theta)^{-1/2}N\left(0, I_N\right) + R,$$

as $t \to \infty$. This suggests the notion of asymptotic efficiency of estimators in terms of the minimal asymptotic variation, when $R = 0$. This justifies the following definition.

Definition 1.2.14. *Assume that the family $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN property at $\theta \in$ int Θ. We say that a family of estimators $\left\{\hat{\theta}(t); t \geq 0\right\}$ is asymptotically efficient at θ in the sense of Hájek-Le Cam convolution theorem if*

$$Law\left(\varphi(t)^{-1}\left(\hat{\theta}(t) - \theta\right)\middle| P_\theta\right) \to I(\theta)^{-1/2}N\left(0, I_N\right) \quad as \quad t \to \infty, \tag{1.3}$$

where I_N denotes the $N \times N-$identity matrix.

We conclude this section by recalling Hajek's minimax theorem Hájek [19] in the context of continuous time stochastic processes.

Theorem 1.2.15 (Minimax Theorem)**.** *Let $\{P_\theta; \theta \in \Theta\}$ be LAMN at any $\theta \in int\, \Theta$ and suppose that $\tilde{\theta}(t)$ is any estimator for θ. Consider $w : \mathbb{R}^N \to [0, \infty)$ a bowl-shaped loss function, that means,*

1. $w(x) \geq 0$;
2. $w(x) = w(-x)$;
3. *the set* $\{x; w(x) \leq c\}$ *is convex in* $\mathbb{R}^N$ *for any* $c > 0$.

Then

$$\lim_{\delta\to 0} \liminf_{t\to\infty} \sup_{|\theta-\theta_0|<\delta} E_{\theta_0} \left[w \left(\varphi^{-1}(t) \left(\tilde{\theta}(t) - \theta_0 \right) \right) \right] \geq E \left[w \left(I(\theta)^{-1/2} Z \right) \right],$$

where $Z \sim N(0, I_N)$ is independent of $I(\theta)$.

Proof. See Ibragimov and Has'minskii [22, Theorem II.11.2 pp.160-161].

Remark 1.2.16.

1. *The Theorem 1.2.15 yields an asymptotic minimax bound from below for an arbitrary loss function;*
2. *If $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN property at θ, defining the class of asymptotically efficient as the estimators which attain the asymptotic minimax bound, then Ibragimov and Has'minskii [22, Theorem II.13.3.] states that this estimators* cannot be superefficient *at this point for any loss function w.*

1.3 Related Literature

This section is dedicated to present some classes of Lévy-driven stochastic differential equations and its recent statistical inference results. For the reader's convenience, the definition of Lévy process is presented below.

Definition 1.3.1 (Lévy process)**.** *Let* **L**$= \{L(t); t \geq 0\}$ *be a stochastic process defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We say that* **L** *is a* Lévy process *if*

L1. $L(0) = 0$ *(a.s.);*

L2. **L** *has independent increments, i.e., $L(t_0), L(t_1) - L(t_0), \cdots L(t_n) - L(t_{n-1})$ are independent random variables for every $0 < t_0 < t_1 < \cdots < t_{n-1} < t_n$ and for all positive integer n;*

L3. **L** *has stationary increments, i.e., for all $t \geq 0$, $L(t+h) - L(t)$ has the same distribution as $L(h)$ for all $h > 0$;*

L4. **L** *is stochastically continuous, i.e., for all $\delta > 0$ and $s \geq 0$*

$$\lim_{t\to s} \mathbb{P}\left(|L(t) - L(s)| > \delta\right) = 0;$$

L5. **L** *has càdlàg paths.*

1.3.1 Langevin Equation

Consider a Lévy process $\mathbf{L} = \{L(t); t \geq 0\}$ on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$. Let $\Theta = \mathbb{R}$ be a parameter space. The *Langevin Equation* is given by

$$\begin{cases} dX(t) &= -\theta X(t)dt + dL(t), \quad t > 0, \\ X(0) &= X_0, \end{cases} \tag{1.4}$$

where the initial solution X_0 is independent of $\mathbf{L}$. The solution of (1.4) is the well-known *Ornstein-Uhlenbeck* process $\mathbf{X} = \{X(t); t \geq 0\}$ which is written as

$$X(t) = X_0 e^{-\theta t} + \int_0^t e^{-\theta(t-s)} dL(s), \quad t \geq 0. \tag{1.5}$$

Küchler and Sϕrensen [33] presented the study of maximum likelihood estimation of the drift parameter when a Wiener process is the noise of (1.4). Using a Lévy process as noise in the Langevin Equation, a non-Gaussian OU-process is obtained as a solution and its important application in finance is well known as a building block of Barndorff-Nielsen and Shephard [3]. For a Lévy $\alpha-$stable noise, Hu and Long [20] proposed an asymptotically consistent least squares estimator which converges to a stable distribution, when it is properly normalized. Mai [42, 43] presented a maximum likelihood estimation study for (1.5).

The parametric statistical experiment associated with (1.5) is $(\mathcal{X} = D[0, \infty), \mathbb{B}(\mathcal{X}), \{P_\theta \in \Theta\})$. Here P_θ is the probability measure induced by $\mathbf{X}$ on $D[0, \infty)$, the space of càdlàg functions on $[0, \infty)$. Under certain conditions (see Mai [42, 43]) the measures $\{P_\theta; \theta \in \Theta\}$ are locally equivalent and the Radon-Nikodym derivative (likelihood function) is given by

$$\frac{dP_\theta^t}{dP_0^t}(X_t) = \exp\left\{ -\frac{\theta}{\sigma^2} \int_0^t X(s) dX^c(s) - \frac{\theta^2}{\sigma^2} \int_0^t X^2(s) ds \right\}, \tag{1.6}$$

where X^c denotes the continuous martingale part of X, $X_t = \{X(s); 0 \leq s \leq t\}$ and P_θ^t denotes the restriction $P_\theta^t = P_\theta|_{\mathcal{F}_t}$.

The maximum likelihood estimator (MLE) of θ is given explicitly by

$$\hat{\theta}(t) = -\frac{\int_0^t X(s) dX^c(s)}{\int_0^t X^2(s) ds}. \tag{1.7}$$

The asymptotic behavior of the MLE $\hat{\theta}(t)$ was studied by Mai [42, 43] who proved (under conditions that ensure ergodicity), for $\theta \in \mathbb{R}_+$, the consistency

$$\hat{\theta}(t) \to \theta \;\; P_\theta - a.s.$$

and the asymptotic normality

$$\text{Law}\left(\sqrt{t}\left(\hat{\theta}(t) - \theta\right) \middle| P_\theta \right) \to N,$$

for a zero mean normal distribution N. Furthermore, he showed that the statistical experiment $\{P_\theta; \theta \in \Theta\}$ is *locally asymptotically normal*, as in Definition 1.2.6. Then, the MLE is asymptotically efficient in the sense of Hajék-Le Cam convolution theorem.

Another problem studied by Mai [42, 43] is how to obtain a discretization of (1.7) since the continuous martingale part X^c is not observed. Inspired by the previous work of Mancini [44], Mai [42, 43] used a thresholding technique to approximate the continuous part of the process. This technique consists of deleting increments that are larger than a threshold $v_n > 0$ and filtering increments that most likely contain jumps. Thus, for a discrete time process $X_{t_1}, \cdots, X_{t_n}$, observed from a path of (1.5), the discretized *filtered MLE* is

$$\bar{\theta}_n := -\frac{\sum_{i=1}^{n} X_{t_i}\Delta_i X \mathbf{1}_{[|\Delta_i X|\le v_n]}}{\sum_{i=1}^{n} X_{t_i}^2 (t_{i+1}-t_i)},$$

where $\Delta_i X = X_{t_{i+1}} - X_{t_i}$. Furthermore, under suitable conditions, he proved asymptotic normality and efficiency.

1.3.2 Stochastic Delay Differential Equation

Consider a Lévy process $\mathbf{L} = \{L(t); t \ge 0\}$ on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \ge 0\}, \mathbb{P})$. Gushchin and Küchler [17] studied necessary and sufficient conditions for the existence of a stationary solution to the *stochastic delay differential equation* (SDDE)

$$\begin{cases} dX(t) &= \int_{[-r,0]} X(t+u)a(du)dt + dL(t), \\ X(t) &= X_0(t), \quad t \in [-r, 0], \end{cases} \tag{1.8}$$

where a is a finite signed measure and $r > 0$ is a fixed number. The solution of (1.8) is uniquely pathwise given by

$$X(t) = x_0(t)X_0(0) + \int_{[-r,0]} \int_u^0 X_0(s)x_0(t+u-s)ds\, a(du) + \int_0^t x_0(t-s)dL(s),$$

where $x_0 : [-r, \infty) \to \mathbb{R}$ is the *fundamental solution*, i.e., $x_0(t) = 0$ for $t \in [-r, 0)$, $x_0(0) = 1$ and

$$x_0'(t) = \int_{[-r,0]} x_0(t+u)a(du)\cdot \tag{1.9}$$

The explicit form of $x_0(t)$ is given in Gushchin and Küchler [17, Lemma 2.1].

The following particular case of (1.8)

$$\begin{cases} dX(t) &= (\theta_1 X(t) + \theta_2 X(t-1))dt + dL(t), \quad t > 0 \\ X(t) &= X_0(t), \quad t \in [-1, 0], \end{cases} \tag{1.10}$$

is obtained by taking $r = 1$ and $a(du) = \theta_1\delta_0(du) + \theta_2\delta_{-1}(du)$, where $\delta_A(du)$ is the Dirac measure and $\theta^\top = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^2$. In this case, (1.9) becomes

$$x_0'(t) = \theta_1 x_0(t) + \theta_2 x_0(t-1)\cdot$$

If $\theta_2 = 0$, then (1.10) is the Langevin equation.

As in the previous subsection, it is known that $\mathbf{X} = \{X(t); t \ge -1\}$, the solution of (1.10), induces a parametric family of measures $\{P_\theta; \theta \in \mathbb{R}^2\}$ on the space of the càdlàg functions $D[-1, \infty)$ and that the likelihood function is

$$\frac{dP_\theta^t}{dP_0^t} = \exp\left\{\theta^\top \begin{pmatrix} \int_0^t X(s)dX^c(s) \\ \int_0^t X(s-1)dX^c(s) \end{pmatrix} - \frac{1}{2}\theta^\top \begin{pmatrix} \int_0^t X^2(s)ds & \int_0^t X(s)X(s-1)ds \\ \int_0^t X(s)X(s-1)ds & \int_0^t X^2(s-1)ds \end{pmatrix} \theta\right\} \tag{1.11}$$

where X^c denotes the continuous $P_\theta-$martingale part (cf. Mai [42]).

Mai [42] showed that for an ergodic version of $\mathbf{X}$ satisfying suitable conditions, the MLE $\hat{\theta}(t)$ of the parameter θ is strongly consistent and asymptotically normal, that is, as $t \to \infty$

$$\hat{\theta}(t) \to \theta \;\; P_\theta - \text{a.s.}$$

and

$$\text{Law}\left(\sqrt{t}\left(\hat{\theta}(t) - \theta\right)\Big| P_\theta\right) \to N,$$

in which N is a $2-$dimensional zero mean normal distribution. Furthermore, observing the solution $\mathbf{X}$ of (1.10) at discrete times $t_1 < \cdots < t_n$, Mai [42] showed that the discretized MLE, using jump filtering as in the Langevin case, is asymptotically normal with the same limit distribution as the continuous time MLE.

Gushchin and Küchler [16] studied (1.10) when it is driven by a Wiener process $\mathbf{W} = \{W(t); t \geq 0\}$. They showed local asymptotic properties of the likelihood function and its strong dependency on the true value of the parameter $\theta^\top = (\theta_1, \theta_2)$. As a consequence, the asymptotic efficiency of the MLE is obtained for the true parameter belonging to some $\Theta_0 \subset \mathbb{R}^2$ (cf. Küchler and Sϕrensen [33, Section 9.4]).

We finalize this subsection highlighting that statistical inference problems for SDDE have been extensively studied in several theoretical aspects (cf. [5, 6, 4, 14, 16, 15, 30, 55, 56]).

1.3.3 Generalized Langevin Equation

The *generalized Langevin equation* (GLE) was proposed by Kubo [29] and Mori [51]. In this *stochastic differential equation* (SDE) the drift is a stochastic process which can depend on the whole process up to the present time t. This SDE is given by

$$\begin{cases} dX(t) &= -\displaystyle\int_0^t X(s)\gamma(t-s)ds + dL(t), \quad t > 0 \\ X(0) &= X_0, \end{cases}$$

where $\mathbf{L} = \{L(t); t \geq 0\}$ is a noise, X_0 is a random variable independent of $\mathbf{L}$ and $\gamma(\cdot)$ is the memory function. The Langevin equation becomes a particular case when we take $\gamma(t) = -\theta\delta_0(t)$.

Assuming that all processes have finite quadratic mean, Kannan [24] studied the solution of GLE. The author showed that any mean square solution process $\mathbf{X} = \{X(t); t \geq 0\}$ of GLE has the form

$$X(t) = X_0(0)\rho(t) + \int_0^t \rho(t-s)dL(s), \quad t > 0, \tag{1.12}$$

where $\rho(\cdot)$ satisfies the Volterra integro-differential equation

$$\begin{cases} \rho'(t) &= -\displaystyle\int_0^t \rho(s)\gamma(t-s)ds, \\ \rho(0) &= 1\cdot \end{cases}$$

This process is called *Generalized Ornstein-Uhlenbeck* (GOU) process.

Santos [59] proved that (1.12) is also a solution for the GLE when $\mathbf{L}$ is an α-stable Lévy process with $1 < \alpha < 2$. Furthermore, Medino et al. [49] showed that (1.12) remains a solution using stochastic integration in the sense of convergence in probability. As mentioned above, the Lévy-driving OU-process can be seen as a particular case of (1.12), then it is natural to be interested in extend estimation results for the GLE.

Stein, Lopes and Medino [62] presented a new approach to the GLE, which generalizes the one previously used. They considered a family of finite signed measures $\{\mu_t; t \geq 0\}$ instead of the memory function $\gamma(\cdot)$. Thus, the GLE becomes

$$\begin{cases} dX(t) &= -\int_0^t X(s)\mu_t(ds)dt + dL(t), \;\; t > 0 \\ X(0) &= X_0, \end{cases}$$

and the function $\rho(\cdot)$ in the Kannan's solution (1.12) satisfies

$$\begin{cases} \rho'(t) &= -\int_0^t \rho(s)\mu_t(ds), \\ \rho(0) &= 1\cdot \end{cases} \tag{1.13}$$

Changing the function $\rho(\cdot)$ in (1.12) changes the self-dependence structure of the process. Furthermore, Stein, Lopes and Medino [62] study the maximum likelihood estimation procedure to estimate the parameters of the discretized process arising from some classes of continuous-time processes that are solution of the GLE. They present results for $\rho(t) = e^{-\theta t}$ (the Ornstein-Uhlenbeck process), $\rho(t) = \cos(\theta t)$ (the Cosine process) and for $\rho(t) = e^{-\theta t^2}$. Below, we summarized some of their results.

Example 1.3.2. *If, for all $\theta > 0$ and each $t \geq 0$, $\mu_t(ds) = \theta\delta_0(t-s)ds$ is a Dirac measure, then $\rho(t) = e^{-\theta t}$ satisfies (1.13). Thus, the obtained process*

$$X_\theta(t) = X_0 e^{-\theta t} + \int_0^t e^{-\theta(t-s)} dL(s)$$

is the well-known OU-process.

Consider discrete times $0, h, 2h, \cdots, kh$ ($k \in \mathbb{N}$ and $h > 0$ fixed) and the process $\mathbf{X}$ observed at this times. It is well known that the OU-process is a stationary autoregressive process satisfying

$$X((k+1)h) = e^{-\theta h} X(kh) + \xi_{k,h},$$

where $\xi_{k,h} = \int_{kh}^{(k+1)h} e^{-\theta((k+1)h-s} dL(s)$.

Example 1.3.3 (see Proposition 4.1 in Stein, Lopes and Medino [62])**.** *If, for all $\theta > 0$ and each $t \geq 0$, $\mu_{\theta,t}(ds) = \theta^2 ds$, then the solution of the GLE is a process of the form (1.12) with $\rho_\theta(t) = \cos(\theta t)$. This process is called Cosine Process and it is written as*

$$X(t) = X_0 \cos(\theta t) + \int_0^t \cos\left(\theta(t-s)\right) dL(s).$$

If $\mathbf{L}$ is an $\alpha-$stable Lévy process, then one discretization form of this process is given by

$$X((k+1)h) = 2\cos(\theta h) X(kh) - X((k-1)h) + \xi_{k,h},$$

where $\xi_{k,h}$ is an $\alpha-$stable $S_\alpha\left((2\int_0^h |\cos(\theta s)|^\alpha ds)^{1/\alpha}, 0, 0\right)$ random variable.

Example 1.3.4 (see Proposition 4.2 in Stein, Lopes and Medino [62])**.** *Consider $\mu_t(ds) = 2\theta(1 - 2\theta(t-s)^2)ds$, for $\theta > 0$. By solving the differential equation (1.13) we obtain $\rho(t) = e^{-\theta t^2}$. Then, for an $\alpha-$stable Lévy process $\mathbf{L}$, the resulting process is*

$$X(t) = X_0 e^{-\theta t^2} + \int_0^t e^{-\theta(t-s)^2} dL(s)\cdot$$

This process has a discretization form given by

$$X((k+1)h) = e^{-\theta(2k+1)h^2} X(kh) + \xi_{k,h},$$

where $\xi_{k,h}$ *is* $S_\alpha\left(\sigma_\xi^\alpha, 0, 0\right)$ *with*

$$\sigma_\xi^\alpha = \int_0^{kh} e^{-\alpha\theta\left((kh-s)^2+(2k+1)h^2\right)} \left(e^{2\theta sh} - 1\right)^\alpha ds + \int_{kh}^{(k+1)h} e^{-\alpha\theta\left((kh-s)^2-2sh+(2k+1)h^2\right)} ds.$$

Remark 1.3.5. *Note from Langevin equation (1.4) and Examples 1.3.3 and 1.3.4 that all the unknown parameter information* $\theta \in \Theta$ *in the GLE is contained in the signed measure* $\mu_t(ds)$ *(or, in the memory function* γ*). By this reason, we will index by* θ *all the measures* $\mu_{\theta,t}(ds)$ *to make clear this parameter dependence. Similarly, we will index the function* $\rho_\theta(t)$ *in the Kannan's solution (1.12).*

Alcântara [1] was interested in modeling the *cupom cambial* time series. However, the difficulty in modeling the *cupom cambial* via classic OU-process is that this time series presents characteristics of an order grater than 1 autoregressive model. Based on this, the author proposed a new class of models satisfying (1.12), when

$$\rho_\theta(t) = (1-\theta_2)e^{-\theta_1 t} + \theta_2 \cos(\theta_1 t), \tag{1.14}$$

$\theta = (\theta_1, \theta_2) \in \mathbb{R}_+ \times [0,1]$, and named *generalized Ornstein-Uhlenbeck of the fluctuating exponential type* process (GOU-FE). He also showed that the proposed process, besides being a solution for the GLE, can be used in the statistical modeling of the *cupom cambial*.

Example 1.3.6 (see Alcântara [1])**.** *Consider* $\theta^\top = (\theta_1, \theta_2) \in \Theta = \mathbb{R}_+ \times [0,1]$ *and* $t \geq 0$*. Let* $\mu_{\theta,t}$ *be a signed measures on* $[0,t]$ *satisfying (1.13) which has the following decomposition*

$$\mu_{\theta,t}(ds) = \theta_1(1-\theta_2)\mu_{\theta,t}^{(1)}(ds) + g_\theta(t-s)ds, \tag{1.15}$$

where $\mu_{\theta,t}^{(1)}(ds) = \delta_0(t-s)$ *is a Dirac measure and* $g_\theta(t)$ *is given by*

$$g_\theta(t) = \begin{cases} e^{-\theta_1\theta_2 t/2}\left(\alpha_1 \cos(\theta_1\nu_1 t) + \alpha_{2,1}\sin(\theta_1\nu_1 t)\right), & \nu_0 > 0, \\ e^{-\theta_1\theta_2 t/2}\left(\alpha_1 \cosh(\theta_1\nu_1 t) + \alpha_{2,-1}\sinh(\theta_1\nu_1 t)\right), & \nu_0 < 0, \\ e^{-\theta_1\theta_2 t/2}\left(\alpha_1 + \alpha_{2,0} t\right), & \nu_0 = 0, \end{cases} \tag{1.16}$$

for constants (functions of θ*) given explicitly by* $\nu_0 = 1 - \theta_2 - \theta_2^2/4$, $\nu_1 = \sqrt{|\nu_0|}$, $\alpha_1 = \theta_1^2\theta_2^2$,

$$\begin{aligned} \alpha_{2,1} &= -\theta_1^2\left[-\frac{1}{\nu_1} - \frac{\theta_2^2}{2\nu_1} + (1-\theta_2)\left(\nu_1 - \frac{\theta_2^2}{4\nu_1}\right)\right], \\ \alpha_{2,-1} &= -\theta_1^2\left[-\frac{1}{\nu_1} + \frac{\theta_2^2}{2\nu_1} + (1-\theta_2)\left(-\nu_1 - \frac{\theta_2^2}{4\nu_1}\right)\right], \\ \alpha_{2,0} &= \theta_1^3\left(\frac{(1-\theta_2)\theta_2^2}{4} + \frac{\theta_2^2}{2} + 1\right). \end{aligned}$$

Alcântara [1] also showed that the process $\mathbf{X}$ *satisfying (1.12) has* $\rho_\theta(t)$ *given by (1.14). Furthermore, the process has a discretization form*

$$X((k+1)h) = \kappa_1(\theta)X(kh) + \kappa_2(\theta)X((k-1)h) + \xi_{k,h}$$

for an $\alpha-$*stable random variable* $\xi_{k,h}$*.*

Another well-studied class of SDE (see [5, 6, 16, 17, 42]) is the class of processes that are the solution of a SDDE. We summarize some inferential results for this process in the Subsection 1.3.2. Now, observe that the drift in (1.10) can be rewritten as

$$\begin{aligned} \theta_1 X(t) + \theta_2 X(t-1) &= -\int_0^t X(s)(-\theta_1 \delta_0(t-s))ds - \int_0^t X(s)(-\theta_2 \delta_0(t-1-s))ds \\ &= -\int_0^t X(s)\mu_{\theta,t}(ds), \end{aligned} \tag{1.17}$$

where $\theta \in \Theta \subset \mathbb{R}^2$ and $\mu_{\theta,t}(ds) = -\theta_1\delta_t(ds) - \theta_2\delta_{t-1}(ds)$ is a linear combination of two Dirac measures. Motivated by these arguments, we finish this subsection presenting formally the notion of GLE.

Definition 1.3.7. *Let* $\mathbf{L} = \{L(t); t \geq 0\}$ *be a Lévy process. Consider* $\mathbf{X}_0 = \{X_0(t); t_0 \leq t \leq 0\}$ *an initial solution process with càdlàg trajectories and independent of* $\mathbf{L}$*. We write the Generalized Langevin Equation as*

$$\begin{cases} dX(t) = b(\theta, X_t)dt + dL(t), & t > 0 \\ X(t) = X_0(t), & t \in [t_0, 0] \end{cases} \tag{1.18}$$

where $X_t = \{X(s); 0 \leq s \leq t\}$ *and* $b(\theta, X_t)$ *is defined as*

$$b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds), \tag{1.19}$$

where $\mu_{\theta,t}$ *is a signed measure on* $[0, t]$*, for each* $t > 0$ *fixed and each parameter* $\theta \in \Theta \subset \mathbb{R}^N$.

Note that, if the family $\{\mu_{\theta,t}; t \geq 0\}$ is such that $b(\theta, X_t) = b(\theta, X(t))$ (in other words, process b depends on $\mathbf{X}$ only at time t), then we can use Applebaum [2, Theorem 6.2.9] to ensure that there exists a unique càdlàg adapted solution to (1.18). But, this Theorem cannot be applied if $b(\theta, \cdot)$ depends on all information from $\mathbf{X}$ until time t. Based on the GLE solution existence studies presented in [24, 49, 59], we present the following definition.

Definition 1.3.8. *Let* $\mathbf{X} = \{X(t); t \geq t_0\}$ *be a stochastic process and* $\rho_\theta(\cdot)$ *be a deterministic function, for each* $\theta \in \Theta$*. We say that the pair* $(\mathbf{X}, \rho_\theta)$ *represents a class of solutions to the Generalized Langevin Equation (1.18) if* $\mathbf{X}$ *is given by*

$$X(t) = X_0(0)\rho_\theta(t) + \int_0^t \rho_\theta(t-s)dL(s), \quad t > 0, \tag{1.20}$$

and $\rho_\theta(\cdot)$ *satisfies the integral equation (1.13), where* $\{\mu_{\theta,t}; t \geq 0$ *and* $\theta \in \Theta\}$ *is a family of signed measures,* $\mathbf{L} = \{L(t); t \geq 0\}$ *is a Lévy process and* $X_0(t)$*,* $t \in [t_0, 0]$*, is the initial process assumed to be independent of* $\mathbf{L}$*. The stochastic process* $\mathbf{X}$ *is called Generalized Ornstein-Uhlenbeck Process.*

The problem of finding an explicit solution for some variations of the GLE has been extensively studied. We have already mentioned [1, 49, 59, 61, 62] who studied the Kannan's solution (1.20) (cf. Kannan [24]) for some cases of GLEs. Fox [12] obtained, via Laplace transform, for a GLE with a Gaussian noise, an explicit solution very similar to Kannan's solution (1.20). Moreover, he discussed about Fokker-Planck-like equations and how they can lead one to the erroneous conclusion that the process is non-stationary, Gaussian and Markovian. McKinley and Nguyen [48], observing that the relationship between the mean-squared displacement (MSD) and the memory structure of the GLE have never been fully characterized (in particular, the special cases where explicit solutions exist), they established a class of memory kernels for which the GLE is well-defined, investigated the associated regularity properties of solutions and proved that large-time asymptotic behavior of the particle MSD is entirely determined by the tail behavior of the GLE's memory kernel. Slezak [60] studied the stationary solutions of the Langevin equation and the GLE. He showed in Slezak [60, Proposition 3.4.1] that if there exists a stationary solution for a GLE, then it has a spectral representation in terms of Fourier transform of the kernel and the spectral representation of the noise. Di Terlizzi, Ritort and Baiesi [9] also studied explicit solutions for a class of GLEs. They introduced a generalisation of the Laplace transform as a useful tool for solving this problem.

Other topics of GLE studies also arise from several types of phenomena (physical, biological, economic, among others). Zhu and Venturi [66] presented a new method to approximate the Mori–Zwanzig memory integral in a class of GLEs for systems with local interactions. They showed that the proposed method is effective in computing autocorrelation functions. Pavliotis, Stoltz and Vaes [53] worked with a simple quasi-Markovian GLE, that is, the GLE is equivalent to a finite-dimensional system of Markovian SDEs. They studied the longtime behavior of solutions and scaling limits of the effective diffusion coefficient associated with the dynamics. Lim, Wehr and Lewenstein [37] studied homogenization for a class of GLEs with state-dependent coefficients and exhibiting multiple time scales.

1.3.4 Other Class of Continuous Time Stochastic Process

Consider a compact subset $\Theta \subset \mathbb{R}^N$. Gloter, Loukianova and Mai [13] studied statistical inference for the parameter $\theta \in \Theta$ based on an observed path $\mathbf{X}_\theta = \mathbf{X} = \{X(t); t \geq 0\}$ which is solution of the integral equation

$$X(t) = X(0) + \int_0^t b(\theta, X(s))ds + \int_0^t \sigma(X(s))dW(s) + \int_0^t \int_{\mathbb{R}} \gamma(X(s^-))zN(ds, dz), \qquad (1.21)$$

where $W(t)$ is a one-dimensional Brownian motion and $N(.,.)$ is the Poisson random measure on $\mathbb{R}_+ \times \mathbb{R}$ associated with the jumps of the Lévy process.

The parametric statistical experiment associated with (1.21) is $(\mathcal{X} = D[0,\infty), \mathbb{B}(\mathcal{X}), \{P_\theta \in \Theta\})$ in which P_θ is the probability measure induced by $\mathbf{X}_\theta$ on $D[0,\infty)$. Gloter, Loukianova and Mai [13] defined the log-likelihood function as

$$l(\theta, X_t) = \int_0^t \frac{b(\theta, X(s))}{\sigma^2(X(s))} dX^c(s) - \frac{1}{2}\int_0^t \frac{b^2(\theta, X(s))}{\sigma^2(X(s))} ds. \qquad (1.22)$$

The choice for the log-likelihood differs from the logarithm of Radon–Nicodym derivative by the multiplicative factor not depending on θ.

Under conditions on existence of ergodic solution and Hölder continuity of the drift $b(\theta, X(t))$, the gradient $\nabla_\theta b(\theta, X(s))$ and the Hessian $\partial_\theta^2 b(\theta, X(s))$, the authors showed that the MLE $\hat{\theta}(t)$ of θ also maximizes $l(\theta, X_t)$, that is

$$\hat{\theta}(t) := \hat{\theta}(X_t) \in \arg\max_{\theta \in \Theta} l(\theta, X_t)\cdot$$

Observe that, unlike the Langevin equation or the SDDE, this MLE is not given explicitly. Then, in order to study the asymptotic behaviour of this estimator, a uniform version of the law of large numbers for martingales is needed. Thus, Gloter, Loukianova and Mai [13] proved the consistency of the MLE $P_\theta-$a.s

$$\hat{\theta}(t) \to \theta \text{ as } t \to \infty$$

and the asymptotic normality

$$\text{Law}\left(\sqrt{t}\left(\hat{\theta}(t)-\theta\right)\Big| P_\theta\right) \to N(0, I^{-1}(\theta)), \text{ as } t\to\infty, \tag{1.23}$$

where $I(\theta)$ is the asymptotic Fisher information. Furthermore, they proved that the statistical experiment $\{P_\theta; \theta \in \Theta\}$ satisfies the LAN property at all $\theta \in \text{int}\,\Theta$. Consequently, the MLE is asymptotically efficient in the sense of Hájek-Le Cam convolution theorem.

Based on discrete high-frequency observations $X(t_0), \cdots, X(t_n)$ of the process (1.21), Gloter, Loukianova and Mai [13] constructed an efficient and asymptotically normal estimator for the drift parameter with minimal conditions on the jumps behaviour and in the way how $\Delta_n = \max\{t_i - t_{i-1}; 1 \leq i \leq n\}$ converges to zero. An important problem to be considered is how to obtain a discretization of $\int_0^t b(\theta, X(s))\sigma^{-2}(X(s))dX^c(s)$ in (1.22). For this, they used a jump filtering technique to propose the following filtered MLE (FMLE)

$$\begin{aligned}\hat{\theta}_n &\in \arg\max_{\theta\in\Theta}\left[\sum_{i=1}^n \sigma(X(t_i))^{-2} b(\theta, X(t_{i-1}))\Delta_i X \mathbf{1}_{|\Delta_i X|\leq v_n^i}\right.\\ &- \left.\frac{1}{2}\sum_{i=1}^n \sigma(X(t_i))^{-2} b(\theta, X(t_{i-1}))^2 (t_i - t_{i-1})\right],\end{aligned}$$

where v_n^i is the cut-off sequence depending on Δ_n and the past information $X(t_1), \cdots, X(t_{i-1})$.

The Theorem 3.2 in Gloter, Loukianova and Mai [13] proves that, under suitable conditions, the FMLE $\hat{\theta}_n$ has the same limit distribution as the continuous time MLE in (1.23), that is

$$\text{Law}\left(\sqrt{t_n}\left(\hat{\theta}_n-\theta\right)\Big| P_\theta\right) \to N(0, I^{-1}(\theta)), \text{ as } t_n\to\infty\cdot$$

1.4 Addressed Issues

Below, we present the main addressed issues in the remaining chapters of this thesis.

1. We extend the studies on estimators for the Langevin equation and SDDE. More precisely, we study MLE for the drift parameter of the GLE observed continuously in time. We show that with appropriated convergence assumptions we have a consistent and asymptotically normal MLE which is also efficient in the sense of Hájek-Le Cam convolution theorem.

2. A filtered MLE is proposed for cases where the GLE is observed on discrete times. We evaluate the results of this discretization via studies of simulations of the GOU-FE process.

3. A new version of discrete time estimator for GLE is proposed.

4. In order to analyze the behavior of the new version of the estimator via simulations, a new class of solution for the GLE is proposed. The process solution of this new class extend the GOU-FE.

5. The two versions of proposed discrete time estimators have so far been studied only through simulations. A more rigorous study of theoretical properties is done for a particular case of GLE. Firstly, the continuous time properties are derived for this particular case of GLE, but with less restriction convergence assumptions than topic 1.

Issues 1 and 2 are studied in Chapter 2. Issues 3 and 4 are covered in Chapter 3. Finally, Chapter 4 address issue 5.

Chapter 2

MLE for the Drift Parameter of the GLE

2.1 Introduction

In this chapter we study MLE for the drift process $b(\theta, X_t)$ associated with the solution process of a Lévy-driven GLE observed continuously in time

$$\begin{cases} dX(t) &= b(\theta, X_t)dt + dL(t), \ \ t > 0 \\ X(t) &= X_0(t), \ \ t \in [t_0, 0] \end{cases} \tag{2.1}$$

where $\{L(t); t \geq 0\}$ is a Lévy noise, $X_t = \{X(s); 0 \leq s \leq t\}$ and $b(\theta, X_t)$ is defined by

$$b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds), \tag{2.2}$$

in which $\mu_{\theta,t}$ a signed measure on $[0, t]$, for each $t > 0$ fixed and each parameter $\theta \in \Theta \subset \mathbb{R}^N$.

In general, the MLE of θ does not possess a closed and explicit form, yet our Theorem 2.3.4 provides sufficient conditions for its consistency, asymptotic normality and efficiency. In particular, we show that the statistical experiment associated with the GLE satisfies the locally asymptotic normal property (LAN as in Definition 1.2.6). A discretization of the MLE is proposed by filtering the "big" jumps and simulation results for FMLE (Filtered MLE) of the generalized Ornstein-Uhlenbeck process of the fluctuating exponential type (GOU-FE) are presented.

Estimation for GOU-FE type processes have been studied by other authors. Alcântara [1] derived a recurrence formula to express a two parameter GOU-FE and obtained estimations via methods used in autoregressive moving average (ARMA) models. Also, for particular cases of the solution process

$$X(t) = X_0(0)\rho_\theta(t) + \int_0^t \rho_\theta(t-s)dL(s), \ \ t > 0, \tag{2.3}$$

estimation results can be found in Alcântara [1], Stein [61] and Stein, Lopes and Medino [62]. More specifically, the processes given by Examples 1.3.3 and 1.3.6. We study the estimation issues under a more general setting by considering the GLE.

In Section 2.2 we describe the statistical experiment associated with the GLE. For the probability measures $\{P_\theta; \theta \in \Theta\}$ associated with the class of solutions $\mathbf{X} = \{\mathbf{X}_\theta; \theta \in \Theta\}$ of (2.1) we explore the likelihood functions (Radon-Nikodym density process of P_θ with respect to $P_{\theta'}$) to obtain some needed proprieties. Propositions 2.2.2 and 2.2.4 give locally absolutely continuous ($\overset{loc}{\ll}$) results for $\{P_\theta; \theta \in \Theta\}$

and can be viewed as a variant of Girsanov's Theorem for the family of measures induced by the solution process $\mathbf{X}$ (see Appendix Theorem A.4.3).

In Section 2.3 we state the asymptotic proprieties for the MLE of θ_0 (the true value),

$$\hat{\theta}(t) := \arg\max_{\theta\in\Theta_0} l(\theta, X_t),$$

where Θ_0 is a convenient subset of Θ and (see Proposition 2.2.2)

$$l(\theta, X_t) := \log \frac{dP_\theta^t}{dP_0^t}(X_t) = \frac{1}{\sigma^2}\int_0^t b(\theta, X_s)dX^c(s) - \frac{1}{2\sigma^2}\int_0^t b^2(\theta, X_s)ds,$$

in which X^c denotes the continuous martingale part of X. Theorem 2.3.4 provides sufficient conditions for $\hat{\theta}(t)$ to be consistent, asymptotically normal and efficient in the sense of Hájek-Le Cam Convolution Theorem (recall the Definition 1.2.14).

In Section 2.4 we make use of the fact that, under regularity conditions, we have $\{P_\theta; \theta \in \Theta\}$ locally equivalent and the Radon-Nikodym derivative is (2.1). To overcome difficulties that arise from the discretization X^c we make use of FMLE that filters the "big" jumps of the process $\mathbf{X}$ as in [13, 42, 43, 44, 45]. Simulation results of FMLE for some GOU-FE models are exhibited.

The last section is dedicated to the proofs of the results. The main tools used are a version of Girsanov Theorem for semimartingales and properties of Hellinger processes (Jacod and Shiryaev [23, Chapters III and IV]) to obtain the likelihood function and the sufficient condition for absolute continuity.

For the asymptotic behaviour of the MLE, Arzelà-Ascoli Theorem, Multivariate Taylor Expansion, Uniform Law of Large Numbers (Loukianova and Loukianov [40]), Central Limit Theorem for continuous time martingales (Küchler and Sϕrensen [34] and Crimaldi and Pratelli [8]) and Wald's method (Vaart [65, Section 5.2]) are used.

2.2 The Statistical Experiment

Consider that the GLE (2.1) is driven by a Lévy process $\mathbf{L} = \{L(t); t \geq 0\}$ on the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ with characteristic triplet (b, σ^2, ν).

The class of solutions $\{\mathbf{X}_\theta; \theta \in \Theta\}$ of (2.1) induces a family of probability measures $\{P_\theta; \theta \in \Theta\}$ on the space $D[t_0, \infty)$ of càdlàg functions on the interval $[t_0, \infty)$ for some fixed t_0. The *parametric statistical experiment* associated with the GLE (2.1) is the triplet $(\mathcal{X} = D[t_0, \infty), \mathbb{B}(\mathcal{X}), \{P_\theta \in \Theta\})$. Here $\mathcal{X}$ is the sample space (the set of observations) and $\mathbb{B}(\mathcal{X})$ is the Borel σ-algebra generated by the Skorohod topology (see Billingsley [7, Chapter 3] or Jacod and Shiryaev [23, Chapter 6] for details).

Consider $\theta_0 \in \Theta$ being the true value of the parameter corresponding to the observed path of $\mathbf{X}$ and let P_θ^t be the restriction $P_\theta^t = P_\theta|_{\mathcal{F}_t}$. As pointed out in (2.1) there is a close connection between the probability measures $\{P_\theta; \theta \in \Theta\}$ and the log-likelihood functions (Radon-Nikodym derivatives).

Understanding the behavior of the Radon-Nikodym derivative of P_{θ_0} with respect to some P_θ will be important to formulate estimates for the parameter θ_0 and study its asymptotic properties. In order to establish the likelihood function for the process $\mathbf{X}$, solution of the GLE (2.1), it is necessary to make some technical hypotheses for the drift process.

Assumption 2.2.1. *We assume that*

1. *for each $\theta \in \Theta$, $b(\theta, X_t)$ is a predictable process (Appendix Definition A.1.6);*

2. $\int_0^t b(\theta, X_s)ds$ *is a finite variation process.*

Now, we establish a version of Girsanov's Theorem for the family of measures induced by $\mathbf{X}$. This result will be proved in Section 2.5 and its proof essentially uses the same arguments as in Proposition 3.2.4 from Mai [42].

Proposition 2.2.2. *Under Assumption 2.2.1, if* $P_\theta \overset{loc}{\ll} P_{\theta'}$ *and the initial measures satisfies* $P_\theta^0 \ll P_{\theta'}^0$, *then the Radon-Nikodym density processes of* P_θ *with respect to* $P_{\theta'}$ *is* $P_{\theta'}$*-a.s. given by*

$$\frac{dP_\theta^t}{dP_{\theta'}^t} = \frac{dP_\theta^0}{dP_{\theta'}^0} \exp\left\{\frac{1}{\sigma^2}\int_0^t \big(b(\theta, X_s) - b(\theta', X_s)\big)\, dX^c(s) - \frac{1}{2\sigma^2}\int_0^t \big(b(\theta, X_s) - b(\theta', X_s)\big)^2 ds\right\}. \tag{2.4}$$

Here X^c *denotes the continuous martingale part of* $\mathbf{X}$ *under* P_θ.

Observe that the likelihood function of the OU process (1.6) and of the SDDE (1.11) become particular cases of (2.4).

The next assumption will gives us sufficient conditions for $P_\theta \overset{loc}{\ll} P_{\theta'}$. Thus, the Proposition 2.2.4 is a generalization of the absolute continuity of the measures induced by the Ornstein-Uhlenbeck processes and can be seen as a variant of Theorem 3.2.1 from Mai [42].

Assumption 2.2.3. *Let* $\{\mu_{\theta,t}; t \geq 0\}$ *from (2.2) be a family of signed measures associated with the function* $\rho_\theta(\cdot)$ *by the Volterra relationship (1.13), for each* $\theta \in \Theta$*. For all* $t \geq 0$ *we assume that*

1. $\rho_\theta(\cdot)$ *is a continuous function on* $[0, t]$*;*

2. $s\int_0^s \mu_{\theta,s}(du) \in L^1([0,t], ds)$*;*

3. $s\int_0^s u\mu_{\theta,s}(du) \in L^1([0,t], ds)$*.*

Note that Assumption 2.2.3 is not particularly restrictive, it holds for Examples 1.3.2, 1.3.3, 1.3.6. Indeed,

(i) If $\rho_\theta(t) = e^{-\theta t}$, $\theta > 0$, $t \geq 0$, then $\mu_{\theta,t}(A) = \theta\delta_A(t)$, where $A \in \mathbb{B}(\mathbb{R}_+)$ and $\delta_A(t)$ is a Dirac measure. Then,

$$s\int_0^s \mu_{\theta,s}(du) \in L^1([0,t], ds). \tag{2.5}$$

Moreover,

$$s\int_0^s u\mu_{\theta,s}(du) \leq s^2\int_0^s \mu_{\theta,s}(du) \in L^1([0,t], ds); \tag{2.6}$$

(ii) If $\rho_\theta(t) = \cos(\theta t)$, $\theta > 0$, $t \geq 0$, then $\mu_{\theta,t}(ds) = \theta^2 ds$. We obtain that

$$s\int_0^s \mu_{\theta,s}(du) = \theta^2 s^2 \in L^1([0,t], ds)$$

and

$$s\int_0^s u\mu_{\theta,s}(du) = \theta^2 s\int_0^s udu \in L^1([0,t], ds);$$

(iii) If $\rho_\theta(t) = (1-\theta_2)e^{-\theta_1 t} + \theta_2 \cos(\theta_1 t)$, $t \geq 0$, $\theta^\top = (\theta_1, \theta_2) \in \mathbb{R}_+ \times [0,1]$, then we have a decomposition of $\mu_{\theta,t}$ as (1.15). By continuity of $g_\theta(\cdot)$ on $[0,t]$ and from the (2.5) and (2.6), we have

$$s\int_0^s \mu_{\theta,s}(du) \quad = \quad s\int_0^s \mu^{(1)}_{\theta,s}(du) + s\int_0^s g_\theta(s-u)du \in L^1\left([0,t], ds\right)$$

and

$$s\int_0^s u\mu_{\theta,s}(du) = (1-\theta_2)\theta_1 s\int_0^s u\mu^{(1)}_{\theta,s}(du) + s\int_0^s u g_\theta(s-u)du \in L^1\left([0,t], ds\right).$$

We finish this section ensuring that Assumption 2.2.3 is a sufficient condition for the absolute continuity of P^t_θ with respect to $P^t_{\theta'}$. We also emphasize that this is the only result of this chapter in which the Kannan's solution (2.3) is used explicitly. For all other results, we require only the differential equation form of the process $\mathbf{X}$ as in (2.1).

Proposition 2.2.4. *Let P_θ and $P_{\theta'}$ be two solution measures for the GOU process (2.3), that is,*

$$X(t) = X_0(0)\rho(t) + \int_0^t \rho(t-s)dL(s), \;\; t > 0.$$

Assume that the initial distributions satisfy $P^0_\theta \ll P^0_{\theta'}$ and that $E_\theta\left(X_0^2\right) < \infty$, for each $\theta \in \Theta$. Then Assumption 2.2.3 is sufficient for $P_\theta \overset{loc}{\ll} P_{\theta'}$.

For the proof in Section 2.5 we make use of the Hellinger process associated with the process $\mathbf{X}$. We briefly review this topic in Section A.5 of the Appendix and refer the reader to Jacod and Shiryaev [23, Chapter IV] for a more detailed reading on the subject.

2.3 The MLE and its Asymptotic Behaviour

Having established the Radon-Nikodym derivatives (2.4) by assuming sufficient conditions on the process $\mathbf{X}$ and on the statistical experiment associated, we are now ready to start the study of statistical inferences for the drift process of the GLE.

Aiming at the addressed issues of statistical inference covered in this chapter, let us assume the general version of the drift process $b(\theta, X_t)$ as defined in (2.2), that is,

$$b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds).$$

In Chapter 4 we will consider $b(\theta, X_t)$ having a particular decomposition form,

$$b(\theta, X_t) = \sum_{j=1}^N \theta_j b_j(X_t), \tag{2.7}$$

where $b_j(X_t) \in L^2\left(dP_\theta \times dt\right)$ and $\theta^\top = (\theta_1, \cdots, \theta_N) \in \Theta \subset \mathbb{R}^N$. It means that all parameter information can be isolated from the process information. In this case, we will see that the MLE has an explicit form and its asymptotic behavior results from the law of large numbers (LLN) and from the central limit theorem (CLT) for martingales. Such approach cannot be used directly when we consider (2.2), since the MLE does not have an explicit form. In this case uniform versions of these results need to be used. For this, further assumptions are required. One must consider a compact subspace of parameter set Θ, Hölder continuity for the drift $b(\theta, X_t)$ and conditions on its gradient $\nabla_\theta b(\theta, X_t)$ and Hessian matrix $\partial_\theta^2 b(\theta, X_t)$.

Now, let $\Theta_0 \ni \theta_0$ be a compact subset of Θ such that $\mathrm{int}\Theta_0$ is non-empty. The Equation (2.4) from Proposition 2.2.2 gives us the log-likelihood function of a solution process for the GLE and it is given by,

$$l(\theta, X_t) := \log \frac{dP_\theta^t}{dP_0^t}(X_t) = \frac{1}{\sigma^2}\int_0^t b(\theta, X_s)dX^c(s) - \frac{1}{2\sigma^2}\int_0^t b^2(\theta, X_s)ds. \tag{2.8}$$

The MLE of θ_0 is given by

$$\hat{\theta}(t) := \arg\max_{\theta\in\Theta_0} l(\theta, X_t). \tag{2.9}$$

Expected properties from $\hat{\theta}(t)$ are consistency (satisfies a Law of Larger Numbers), the distribution of its error (a version of Central Limit Theorem) and some choice criteria to compare it with other estimators. We adopt the criterion of asymptotically efficient in the sense of Hájek-Le Cam (see Definition 1.2.14),

$$\text{Law}\left(\varphi(t)^{-1}\left(\hat{\theta}(t) - \theta\right)\Big| P_\theta\right) \to I(\theta)^{-1/2} N\left(0, I_N\right) \ \text{ as } \ t \to \infty, \tag{2.10}$$

where $I(\theta)$ is the information matrix and I_N the identity matrix. To establish (2.10) one needs to show that the statistical experiment satisfies the LAN property (Definition 1.2.6). Conditions to assure the LAN property are, in general, very complex. Typically one needs ergodicity of the process and a version of *Central Limit Theorem* for continuous-time multivariate martingale. We will make use of CLT for martingales given by Küchler and Sϕrensen [33, 34] (for extension of this result see Crimaldi and Pratelli [8]). As for the ergodicity, Masuda [46, 47] gave a set of conditions under which a multidimensional diffusion with jumps fulfills the ergodic theorem and Kulik [31] gave sufficient conditions for exponential ergodicity of a Markov process defined as the solution to a SDE with jump noise. Other references in this matter include Gloter, Loukianova and Mai [13], Kohatsu-Higa, Nualart and Tran [28], Mai [42, 43], Liu, Nualart and Tindel [39] and Tran [63, 64].

We consider that there exists a positive increasing function $\varphi(t) \uparrow \infty$. In order to get asymptotic results, it is necessary to assume some technical restrictions on $b(\theta, X_t)$, Θ_0 and $\varphi(t)$.

Assumption 2.3.1. *Let $\Theta_0 \ni \theta_0$ be a compact subset of Θ such that intΘ_0 is non-empty and assume that there exists a positive increasing function $\varphi(t) \uparrow \infty$ satisfying,*

1. *for all $\theta \in \Theta_0$, we have*

$$\frac{1}{\sigma^2\varphi(t)}\int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right)^2 ds \to \tilde{\Sigma}^2(\theta), \tag{2.11}$$

P_{θ_0}−a.s. as $t \to \infty$;

2. *the drift process (2.2) is Hölder continuous on θ, i.e. for all $x \in D[0,\infty)$ there exists $K(x_t)$ such that for all $\theta, \theta' \in \Theta_0$ we have*

$$\left|b(\theta, x_t) - b(\theta', x_t)\right| \le K(x_t)\left|\theta - \theta'\right|^\kappa,$$

where $0 < \kappa \le 1$ and $\frac{1}{\varphi(t)}\int_0^t K^2(x_s)ds \to \Sigma_K^2 < \infty$;

Assumption 2.3.2. *Assume that:*

1. *for all $x \in D[0,\infty)$, $b(\cdot, x_t)$ is twice continuous differentiable with respect to θ and $\nabla b(\cdot, x_t)$ and $\partial^2 b(\cdot, x_t)$ are Hölder continuous with respect to θ. That means, for all $\theta, \theta' \in$ int Θ_0,*

$$\left|\nabla b(\theta, x_t) - \nabla b(\theta', x_t)\right| \leq K_1(x_t) \left|\theta - \theta'\right|^{\kappa_1}$$

and

$$\left|\partial^2 b(\theta, x_t) - \partial^2 b(\theta', x_t)\right| \leq K_2(x_t) \left|\theta - \theta'\right|^{\kappa_2},$$

where $0 < \kappa_1, \kappa_2 \leq 1$ and $\frac{1}{\varphi(t)} \int_0^t K_j^2(x_s)ds \to \Sigma_{K_j}^2 < \infty$, for $j = 1, 2$;

2. *for all $j \in \{1, \cdots, N\}$ and each $\theta \in \Theta_0$ we can interchange the following orders of differentiation and stochastic integration*

$$\partial_{\theta_j} \int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right)^2 ds = 2 \int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right) \partial_{\theta_j} b(\theta, X_s) ds$$

and

$$\partial_{\theta_j} \int_0^t b(\theta, X_s) - b(\theta_0, X_s) dW(s) = \int_0^t \partial_{\theta_j} b(\theta, X_s) dW(s),$$

where $\mathbf{W} = \{W(t); t \geq 0\}$ is the Wiener process in the Lévy-Itô decomposition of $\mathbf{L}$;

3. *the asymptotic Fisher Information $I(\theta) = (I_{ij}(\theta))$ and the technical matrix $\xi(\theta) = (\xi_{ij}(\theta))$ (which appear in proof of Theorem 2.3.4) satisfy the limits*

$$\frac{1}{\varphi(t)} \left(\frac{1}{\sigma^2} \int_0^t \partial_{\theta_i} b(\theta, X_s) \partial_{\theta_j} b(\theta, X_s) ds\right)_{1 \leq i,j \leq N} \longrightarrow I(\theta) \tag{2.12}$$

and

$$\frac{1}{\varphi(t)} \left(\frac{1}{\sigma^2} \int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right) \partial^2_{\theta_i \theta_j} b(\theta, X_s) ds\right)_{1 \leq i,j \leq N} \longrightarrow \xi(\theta) \tag{2.13}$$

$P_{\theta_0} - a.s.$ as $t \to \infty$.

See Hutton and Nelson [21] and the references therein for sufficient conditions for the item 2 in Assumption 2.3.2 holds true. The item 3 in the Assumption 2.3.2 has a technical importance for the general case of $b(\theta, X_t)$ because the covariation matrix of the MLE's limit distribution will be $I(\theta_0)^{-1}$.

Magdziarz and Weron [41, Theorem 1] presented an ergodic theorem that can be applied for the GOU process, that is, taking a appropriated Lévy process for which the correlation function of $\mathbf{X}$ decays to zero as $t \to \infty$ and for a suitable f, the temporal and ensemble averages coincide, i.e.,

$$\frac{1}{T} \int_0^T f(X(t))dt \to \mathbf{E}[f],$$

provided that $\mathbf{E}[f] < \infty$. It becomes a sufficient conditions for the convergences assumptions in (2.11), (2.12) and (2.13).

Assumption 2.3.3. 1. *Assume that*

$$E_{\theta_0} \left[\frac{1}{\varphi(t)} \int_0^t K_1^2(X_s) ds\right] \to \Sigma_K^2;$$

2. *for all* $h \in \mathbb{R}^N$ *and* $u, u' \in [0,1]$,

$$\frac{1}{\varphi(t)} \int_0^t \left| \nabla_\theta b(\theta_0, X_s) \nabla_\theta b(\theta_0, X_s)^\top - \nabla_\theta b(\theta_0 + h\varphi(t)^{-1/2}u, X_s) \nabla_\theta b(\theta_0 + h\varphi(t)^{-1/2}u', X_s)^\top \right| ds$$

converges to zero P_{θ_0}*−a.s. as* $t \to \infty$.

Finally, we can state the main result of this chapter.

Theorem 2.3.4. *Under the hypotheses of Proposition 2.2.2 and assuming that the Assumptions 2.3.1, 2.3.2 and 2.3.3 are satisfied we have,*

1. *the MLE (2.9) is strongly consistent and asymptotically normal, that is,*

$$\hat{\theta}(t) \to \theta_0 \;\; P_{\theta_0} - a.s. \;\; as \;\; t \to \infty$$

and

$$Law \left(\varphi(t)^{1/2} \left(\hat{\theta}(t) - \theta_0 \right) \middle| P_{\theta_0} \right) \to N(0, I^{-1}(\theta_0)) \;\; as \;\; t \to \infty;$$

2. *the statistical experiment* $\{P_\theta; \theta \in \Theta_0\}$ *is LAN for each* $\theta \in$ *int* Θ_0 *with the Fisher information matrix* $I(\theta)$ *defined in (2.12) and rate of convergence* $\varphi(t)^{-1/2}$*. That means, for all* $h \in \mathbb{R}^N$,

$$Law \left(\log \frac{dP^t_{\theta + h\varphi(t)^{-1/2}}}{dP^t_\theta} \middle| P_\theta \right) \to h^\top N\left(0, I(\theta)\right) - \frac{1}{2} h^\top I(\theta) h, \;\; as \;\; t \to \infty \cdot$$

Furthermore, the MLE $\left\{ \hat{\theta}(t); t \geq 0 \right\}$ *is asymptotically efficient in the sense of Hájek-Le Cam Convolution Theorem.*

We point out that the Theorem 2.3.4 generalizes results from Mai [42, 43] for the Langevin equation and a class of SDDE. Our proof borrows the ideas from Gloter, Loukianova and Mai [13] and Loukianova and Loukianov [40]. In Chapter 4, we will prove a similar result but for the case when the drift process satisfies (2.7). In this case, less restrictive assumptions will be required.

Using autoregressive decomposition and spectral density for an observed path of the GOU process (2.3) at discrete times, it is possible to use time series modeling for this type of processes (see, for example, [1, 62]). However, as far as we know, there is no studies in the literature on statistical estimation for the drift of a GLE solution process observed continuously in the time interval $[0, t]$. Furthermore, we do not need the explicit form of the Kannan's solution (2.3) to derive Theorem 2.3.4. This is important because, since (2.3) is not necessarily a solution for the GLE, it depends on the sense of integration used and on the conditions over the noise **L** (see (1.10), for example).

Also, Theorem 2.3.4 allows us to model processes with presence of oscillations or seasonality, as occurs in several physical and climatic phenomena, among others. Just make the proper choice of $\mu_{\theta,t}$ or $\rho_\theta(t)$ and modification of the autocorrelation structure of the process arises. In several cases it is possible to obtain processes with oscillatory decay of the autocorrelation function. Oscillations can also occur in the paths of the process itself as in the GOU-FE process when we take $\rho_\theta(t) = (1 - \theta_2)e^{-\theta_1 t} + \theta_2 \cos(\theta_1 t)$.

2.4 The FMLE and Simulation Results

The aim of the this section is to evaluate the applicability of the MLE (2.9) via estimation in simulated processes. We propose a discretization form of $\hat{\theta}(t)$ based on Gloter, Loukianova and Mai [13], Mai [42, 43] and Mancini [44, 45] by filtering big jumps to approximate some stochastic integrals. Then, Monte Carlo simulations are done for the GOU-FE process

$$X(t) = X_0\left((1-\theta_2)e^{-\theta_1 t} + \theta_2\cos(\theta_1 t)\right) + \int_0^t \left((1-\theta_2)e^{-\theta_1(t-s)} + \theta_2\cos(\theta_1(t-s))\right) dL(s) \tag{2.14}$$

in order to answer the following questions: how good the joint estimation of $\theta^\top = (\theta_1, \theta_2)$ is? Does the estimator have the same behaviour in different regions of the plan, contained in $\mathbb{R}_+ \times [0,1]$?

We start this section by setting some notations that were used here and in the figures and table of the simulation results.

Let $f : [0,T] \to \mathbb{R}$ be a real function, with final time $T > 0$. Consider a discretization in $n+1$ steps of the interval $[0,T]$ as $(t_k = (k-1)T/n)_{k=1,\cdots,n+1}$. The increment of f on $[t_i, t_{i+1}]$ is denoted by $\Delta_i f = f(t_{i+1}) - f(t_i)$ and the interval increment is $\Delta_i = T/n$.

We consider a Lévy process $\mathbf{L} = \{L(t); t \geq 0\}$ given by

$$L(t) = \sigma W(t) + \sum_{i=0}^{N(t)} Y_i, \tag{2.15}$$

where $\sigma > 0$ is fixed, $\{W(t); t \geq 0\}$ is a Wiener process and $\{N(t); t \geq 0\}$ a Poisson process with rate $\lambda > 0$ and $\{Y_i\}_{i=0,\cdots,\infty}$ is a sequence of i.i.d. random variables with common distribution $N(0,2)$.

Let $\mathbf{X} = \{X(t); t \geq 0\}$ be a solution process of the GLE (2.1). Denote all the past history of the discretization until the time t_i by $X_{t_i} = \{X(t_j); j \leq i\}$.

The main problem in the discretization of $\hat{\theta}(t)$ defined in (2.9) is getting a good approximation of the increments of the continuous martingale part X^c, because it is not observable. Under P_{θ_0}, (cf. Jacod and Shiryaev [23, Theorem I.4.18 and Proposition I.4.27]) X^c is given by

$$X^c(t) = \sigma W(t) + \int_0^t b(\theta_0, X_s)ds,$$

where $\mathbf{W} = \{W(t); t \geq 0\}$ is the Wiener process in the Lévy-Itô decomposition of $\mathbf{L}$.

In general, the increments $\Delta_i X^c$ are of the order $\Delta_n^{1/2}$ (cf. Mancini [44, 45]). Thus, based on Gloter, Loukianova and Mai [13] and Mai [42, 43], we approximate the discretization of the stochastic integral $\int_0^t b(\theta, X_s)dX^c(s)$ by increments of $\mathbf{X}$ filtering "big" jumps, i.e.

$$\sum_{i=1}^n b(\theta, X_{t_i})\Delta_i X \mathbf{1}_{[|\Delta_i X| \leq v_n^i]},$$

where the cutoff sequence is given by $v_n^i = a_i \Delta_n^\beta$, for $\beta \in (0, 1/2)$ and a_i being a measurable function of the past information of the process. Thus, the *filtered maximum likelihood estimator* (FMLE) is defined as

$$\hat{\theta}_T^{FMLE} := \arg\max_{\theta \in \Theta_0} \left\{ \frac{1}{\sigma^2}\sum_{i=1}^n b(\theta, X_{t_i})\Delta_i X \mathbf{1}_{[|\Delta_i X| \leq v_n^i]} - \frac{1}{2\sigma^2}\sum_{i=1}^n b^2(\theta, X_{t_i})\Delta_i \right\}. \tag{2.16}$$

For the applications in this section, we consider the GOU-FE process which is defined in (2.14). Alcântara [1, Theorem 5.1] stated that, for $\theta^\top = (\theta_1, \theta_2) \in \Theta = \mathbb{R}_+ \times [0,1]$ and $t \geq 0$, the signed measures $\mu_{\theta,t}$ on $[0,t]$ satisfying

$$\begin{cases} \rho_\theta'(t) &= -\displaystyle\int_0^t \rho_\theta(s)\mu_{\theta,t}(ds), \\ \rho_\theta(0) &= 1 \end{cases}$$

have the following decomposition

$$\mu_{\theta,t}(ds) = \theta_1(1-\theta_2)\mu_{\theta,t}^{(1)}(ds) + g_\theta(t-s)ds,$$

where $\mu_{\theta,t}^{(1)}(ds) = \delta_0(t-s)$ is a Dirac measure and $g_\theta(t)$ is given by

$$g_\theta(t) = \begin{cases} e^{-\theta_1\theta_2 t/2}\left(\alpha_1 \cos(\theta_1\nu_1 t) + \alpha_{2,1}\sin(\theta_1\nu_1 t)\right), & \nu_0 > 0, \\ e^{-\theta_1\theta_2 t/2}\left(\alpha_1 \cosh(\theta_1\nu_1 t) + \alpha_{2,-1}\sinh(\theta_1\nu_1 t)\right), & \nu_0 < 0, \\ e^{-\theta_1\theta_2 t/2}\left(\alpha_1 + \alpha_{2,0} t\right), & \nu_0 = 0, \end{cases}$$

for constants (functions of θ) given explicitly by $\nu_0 = 1 - \theta_2 - \theta_2^2/4$, $\nu_1 = \sqrt{|\nu_0|}$, $\alpha_1 = \theta_1^2\theta_2^2$,

$$\begin{aligned} \alpha_{2,1} &= -\theta_1^2\left[-\frac{1}{\nu_1} - \frac{\theta_2^2}{2\nu_1} + (1-\theta_2)\left(\nu_1 - \frac{\theta_2^2}{4\nu_1}\right)\right], \\ \alpha_{2,-1} &= -\theta_1^2\left[-\frac{1}{\nu_1} + \frac{\theta_2^2}{2\nu_1} + (1-\theta_2)\left(-\nu_1 - \frac{\theta_2^2}{4\nu_1}\right)\right], \\ \alpha_{2,0} &= \theta_1^3\left(\frac{(1-\theta_2)\theta_2^2}{4} + \frac{\theta_2^2}{2} + 1\right). \end{aligned}$$

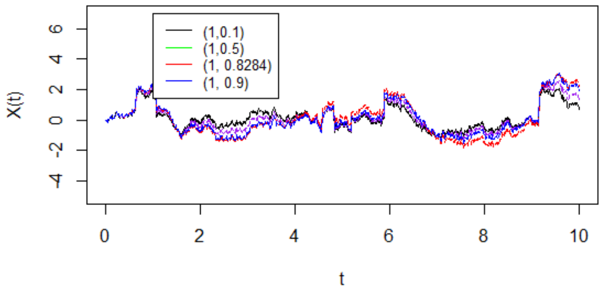
Thus, the parameter space has tree regions in which the process changes its self-dependence that is either $\mathbb{R}_+ \times [0, 2\sqrt{2}-2)$, $\mathbb{R}_+ \times \{2\sqrt{2}-2\}$ or $\mathbb{R}_+ \times (2\sqrt{2}-2, 1]$. Furthermore, the Euler-Maruyama discretization of the GOU-FE were given in Alcântara [1, Equation (5.65)].

Figure 2.1 shows the oscillatory paths of the simulations from GOU-FE processes for $T \in \{10, 50\}$ and for different values of $\theta^\top = (\theta_1, \theta_2)$. As expected, for small θ_2 values, the process has less oscillation. But keeping θ_2 fixed and increasing θ_1, we have more frequency of oscillations.

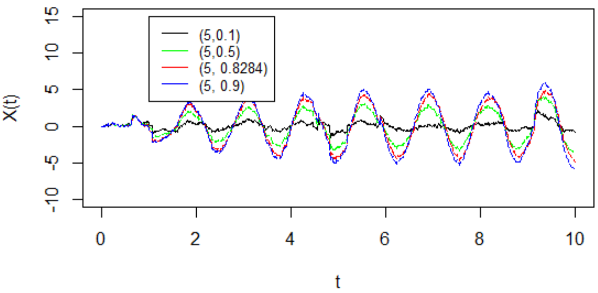
Figure 2.2 shows some simulations paths of the GOU-FE as the corresponding Lévy process and its decomposition (2.15). That means, the left graphics have plotted:

1. the simulated Lévy process (with the legend *lp*) $L(t)$ and its decomposition (2.15);
2. the corresponding Brownian motion (*wp*) $W(t)$;
3. the jump process (*jp*) $J(t) = \sum_{i=0}^{N(t)} Y_i$;
4. the corresponding GOU-FE $X(t)$ for the Lévy path $L(t)$ with the fixed parameter θ.

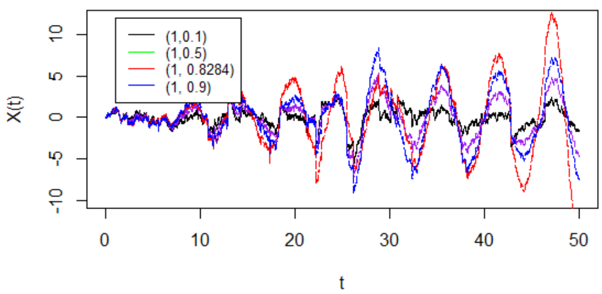
The autocorrelation functions of those GOU-FE are also plotted and have oscillating decays with frequency depending on the θ value. The decay of the autocorrelation function shown in the Figure 2.2 is expected for classes of solution of the GLE, even when considering other GLE variants (see, for example, Alcântara [1, Figures 5.2, 5.4, 5.6, 5.8 and 5.10], Morgado et al. [50, Figure 1], Slezak [60, Figure 6.6] and Zhu and Venturi [66, Figures 3 and 4]). That kind of decay of the autocorrelation function indicates that these processes are autoregressive with order greater than 1. In Chapter 3 we explore the behaviour of the autocorrelation function proving that a 3-parameter GOU-FE has an order 3 autoregressive decomposition.



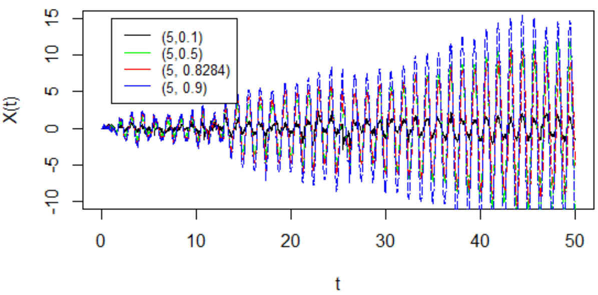
(a) $\theta_1 = 1, T = 10, n = 1000$



(b) $\theta_1 = 5, T = 10, n = 1000$

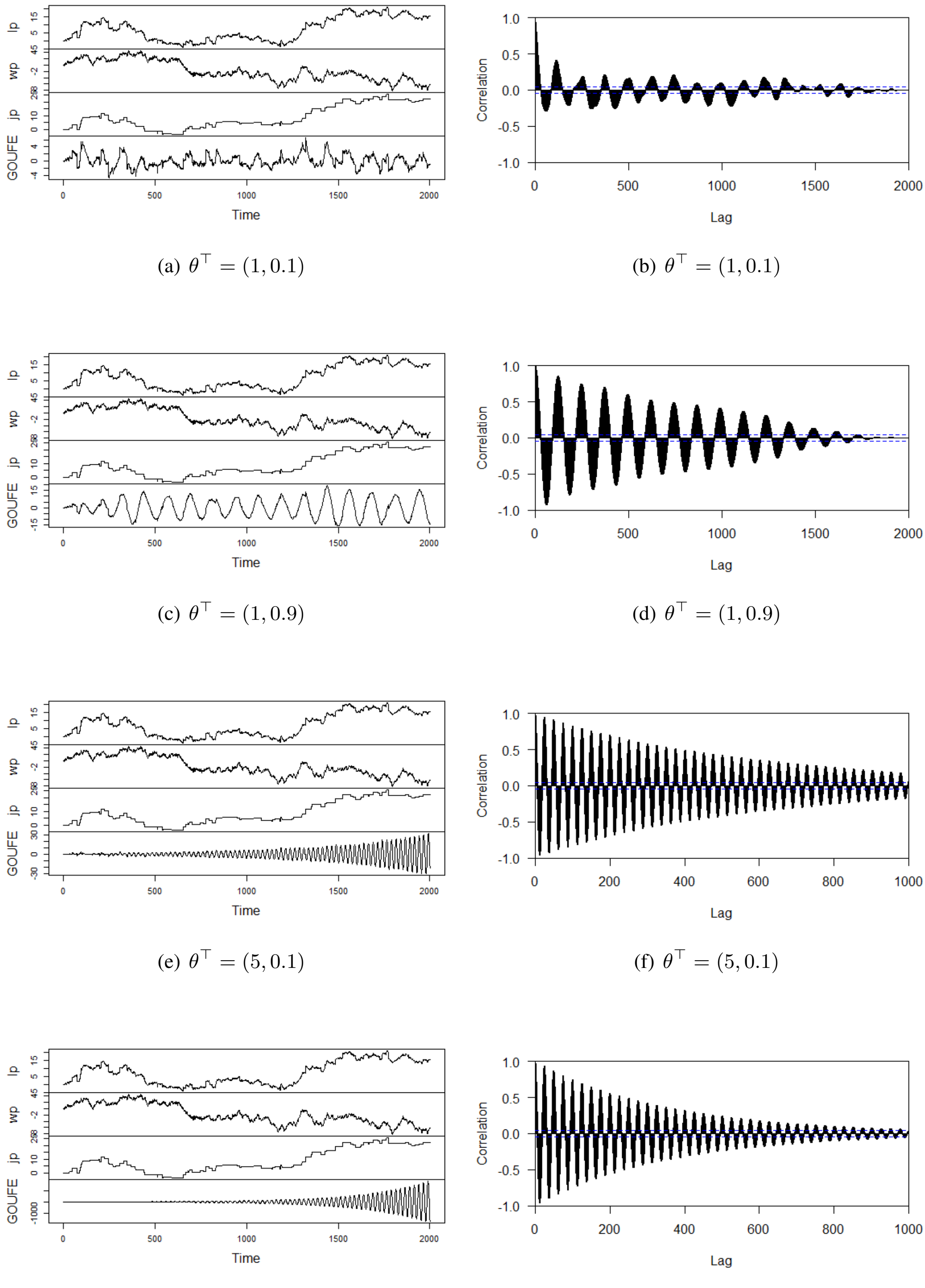


(c) $\theta_1 = 1, T = 50, n = 5000$



(d) $\theta_1 = 5, T = 50, n = 5000$

Figure 2.1: Simulation of the GOU-FE process with fixed values $\sigma = 1$, $\lambda = 1$, $n = 1000$ and $T \in \{10, 50\}$ for different values of $\theta^\top = (\theta_1, \theta_2)$.



(a) $\theta^\top = (1, 0.1)$ (b) $\theta^\top = (1, 0.1)$

(c) $\theta^\top = (1, 0.9)$ (d) $\theta^\top = (1, 0.9)$

(e) $\theta^\top = (5, 0.1)$ (f) $\theta^\top = (5, 0.1)$

(g) $\theta^\top = (5, 0.9)$ (h) $\theta^\top = (5, 0.9)$

Figure 2.2: Simulations (left) and autocorrelation function (right) of the GOU-FE processes with corresponding Lévy process (lp) and its Wiener (wp) and jump (jp) parts as in (2.15), for fixed values $\sigma = 1, \lambda = 1$, $T = 100$, $n = 2000$ and $\theta \in \{1, 5\} \times \{0.1, 0.9\}$.

Table 2.1: Mean, standard deviation and absolute error of the FMLE $\hat{\theta}_{100}^{FMLE}$ for 500 Monte Carlo simulations of the GOU-FE with $n = 1000$ and $\sigma = 1$.

λ	θ_1	θ_2	mean$(\hat{\theta}_{1,T})$	std dv$(\hat{\theta}_{1,T})$	$\lvert\hat{\theta}_{1,T} - \theta_1\rvert$	mean$(\hat{\theta}_{2,T})$	std dv$(\hat{\theta}_{2,T})$	$\lvert\hat{\theta}_{2,T} - \theta_2\rvert$
1	1	0,1000	1,0062	0,0284	0,0062	0,1012	0,0347	0,0012
5	1	0,1000	1,0077	0,0236	0,0077	0,1047	0,0346	0,0047
1	5	0,1000	4,7340	0,0690	0,2660	0,1299	0,0109	0,0299
5	5	0,1000	4,6853	0,0632	0,3147	0,1293	0,0101	0,0293
1	1	0,3000	1,0007	0,0048	0,0007	0,2195	0,0344	0,0805
5	1	0,3000	1,0010	0,0069	0,0010	0,2233	0,0399	0,0767
1	5	0,3000	3,0000	0,0000	2,0000	0,3186	0,0055	0,0186
5	5	0,3000	3,0000	0,0000	2,0000	0,3187	0,0049	0,0187
1	1	0,8284	1,0000	0,0000	0,0000	0,2794	0,0768	0,5490
5	1	0,8284	1,0000	0,0000	0,0000	0,2454	0,0695	0,5830
1	5	0,8284	3,0000	0,0000	2,0000	0,3208	0,0038	0,5077
5	5	0,8284	3,0000	0,0000	2,0000	0,3202	0,0018	0,5083
1	1	0,9000	1,0004	0,0037	0,0004	0,2990	0,0885	0,6010
5	1	0,9000	1,0000	0,0000	0,0000	0,2588	0,0743	0,6412
1	5	0,9000	3,0000	0,0000	2,0000	0,3408	0,0052	0,5592
5	5	0,9000	3,0000	0,0000	2,0000	0,3390	0,0059	0,5610

Table 2.1 is obtained following the step-by-step bellow: For each line in the table

1. columns 1-3 present the fixed parameters of the simulations, in which λ is the rate of the Poisson process $N(t)$ in the decomposition of the Lévy (2.15) and θ_1 and θ_2 are the drift parameters of the GLE;
2. fix a search region Θ_0 for the FMLE;
3. it is simulated 500 Lévy processes;
4. it is obtained the 500 corresponding GOU-FE processes;
5. for each simulation, the parameter $\theta^\top = (\theta_1, \theta_2)$ is estimated, according to the FMLE (2.16);
6. the remaining columns (4-9) present the mean, standard deviation (std dv) and absolute deviation of the estimated parameters.

The results obtained in Table 2.1 were programmed in the language and environment for statistical computing *R* and compiled on a server with 32 cores, in addition to using parallel computing to generate the 500 samples of the parameter settings described in each row of the table. The average computing time was 17 hours per line.

Table 2.1 shows that the rate of jumps λ has no influence on the θ_0 estimation. Also, $\hat{\theta}_{100}^{FMLE}$ estimates well for small θ_2, i.e., good estimates occurred only in the a subset of the first region $\Theta_0 = \mathbb{R}_+ \times [0, 2\sqrt{2} - 2)$. To make the estimator applicable to $\mathbb{R}_+ \times \{2\sqrt{2} - 2\}$ and $\mathbb{R}_+ \times [2\sqrt{2} - 2, 1]$, more sophisticated methods of optimization are required, such as introducing the dependence on a simulated Lévy path or techniques such as VNS (Variable Neighborhood Search).

For θ_2 values greater than 0.1, Table 2.1 shows some standard deviations equal to zero, in some estimations of θ_1. That means, the local search algorithm used has reached the minimum of the searched region. If we took a larger search region, the algorithm would stabilize the estimation around some value from the initially estimated. Even so, the estimates when $\theta_1 = 5$ would not be good, even increasing the search region. This indicates that for large values of θ_1, we may not be able to verify some of the convergences assumed in the previous section, when de perturbation parameter θ_2 is greater than 0.1.

In Chapter 3, we propose a new discrete time estimator for θ_0 in order to improve the estimation for "large" values of θ_1. For this, we introduce the Lévy path's information in the estimator, since in (2.16) we used only σ^2 which is associated with the the Gaussian part of the Lévy process.

2.5 Proofs

Our proofs essentially join and repeat the arguments from [13, 40, 42, 43], adapted to our needs. We refer to Sections A.1, A.2, A.3, A.4 and A.5 for the technical terms that have not been defined so far in this chapter and are used in proofs of Propositions 2.2.2 and 2.2.4.

Proof of the Proposition 2.2.2

Proof of the Proposition 2.2.2. Let $\mathbf{L} = \{L(t); t \geq 0\}$ be a Lévy process with characteristic triplet (b, σ^2, ν). It is well known that the semimartingale characteristic of $\mathbf{L}$ is given by $(bt, \sigma^2 t, \nu(dx)dt)$, where dt denotes the Lebesgue measure. When $\mathbf{X} = \{X(t); t \geq 0\}$ is solution of the GLE, by the Lévy-Itô decomposition we have

$$\begin{aligned} X(t) &= \int_0^t b(\theta, X_s)ds + L(t) \\ &= \int_0^t b(\theta, X_s)ds + bt + \sigma W(t) + \int_{|x|\geq 1} xN(t, dx) + \int_{|x|<1} x\tilde{N}(t, dx), \end{aligned}$$

where $N(t, dx)$ is a Poisson random measure associated with the jumps of the Lévy process and $\tilde{N}(t, dx)$ its compensated Poisson random measures.

Under the Assumption 2.2.1, $\int_0^t b(\theta, X_s)ds$ is of finite variation. Thus, it follows from Applebaum [2, Section 2.7] that the semimartingale characteristic of $\mathbf{X}$ is given by

$$\left(bt + \int_0^t b(\theta, X_s)ds, \sigma^2 t, \nu(dx)dt \right). \tag{2.17}$$

Let (B, C, ν) and (B', C', ν') be the semimartingale characteristics of $\mathbf{X}$ under P_θ and $P_{\theta'}$, respectively. We have that

$$B = bt + \int_0^t b(\theta, X_s)ds \text{ and } B' = bt + \int_0^t b(\theta', X_s)ds.$$

It follows that

$$B = B' + \int_0^t \left(b(\theta, X_s) - b(\theta', X_s) \right) ds.$$

We define the processes

$$\beta(t) = \beta(\theta, \theta', X_t) := b(\theta, X_t) - b(\theta', X_t)$$

and

$$K(t) = K(\theta, \theta', X_t) := \int_0^t \left(\frac{\beta(s)}{\sigma} \right)^2 ds.$$

Since $b(\theta, X_t)$ is a predictable process for each $\theta \in \Theta$ and $\beta^2(t)$ is a non-decreasing process, we conclude that $K(t)$ is a predictable process and has non-decreasing paths. Moreover, $\{K(t); t \geq 0\}$ does not jumps to infinity.

Define a sequence of stopping times $\{T_n; n \in \mathbb{N}\}$ by

$$T_n = \inf\{t \in \mathbb{R}; K(t) \geq n\}$$

and let $A = \bigcup_n [0, T_n]$ be a predictable set. Note that $K(t) \to \infty$ a.s. under P_θ and $P_{\theta'}$ as $t \to \infty$, which implies that

$$A = \mathbb{R}_+ \tag{2.18}$$

P_θ- and $P_{\theta'}$-almost surely.

Let $\tilde{\nu}$ denotes the compensator of ν on $\tilde{\Omega} = \Omega \times \mathbb{R}_+ \times \mathbb{R}$. As in (A.2), by Jacod and Shiryaev [23, Proposition II.1.17], there exists a version of $\tilde{\nu}$ such that $\tilde{\nu}(\omega; \{t\} \times \mathbb{R}) \leq 1$. We define a predictable process $\mathbf{a} = \{a(t); t \geq 0\}$ by

$$a(\omega, t) := \tilde{\nu}(\omega; \{t\} \times \mathbb{R}).$$

For each measurable function Y on $\tilde{\Omega}$ we define

$$\hat{Y}(\omega, t) := \begin{cases} \displaystyle\int_{\mathbb{R}} Y(\omega, t, x)\tilde{\nu}(\omega; \{t\} \times dx) & \text{if this integral converges,} \\ +\infty & \text{otherwise.} \end{cases}$$

According to Girsanov's Theorem A.2.5 (cf. Jacod and Shiryaev [23, Theorem III.3.24]), we have $\tilde{\nu}'(\omega; \{t\} \times \mathbb{R}) = \int_{\mathbb{R}} Y(\omega, t, x)\tilde{\nu}(\omega; \{t\} \times dx)$. It follows from semimartingale characteristics of $\mathbf{X}$ that $Y \equiv 1$ and $\hat{Y}(t) = \tilde{\nu}(\omega; dt \times \mathbb{R}) \stackrel{def}{=} a(\omega, t)$. Thus, given a stopping time S, we obtain

$$\left(Y - 1 - \frac{\hat{Y} - a}{1 - a} \mathbf{1}_{[a<1]} \right) \mathbf{1}_{[0,S]} * (\nu - \tilde{\nu}) \equiv 0.$$

We concluded from Proposition A.2.7 (cf. Jacod and Shiryaev [23, Proposition III.5.10]) that there exists a process $\mathbf{U} = \{U(t); t \geq 0\}$ such that for all stopping time S the stopped process $U(S)$ is given by

$$U(S) = \int_0^S \beta(s) dX^c(s).$$

Thus, $\mathbf{U}$ is a continuous process and

$$\prod_{s<t} (1 + \Delta U(s))\, e^{-\Delta U(s)} \equiv 1,$$

for all $t \geq 0$.

By applying Theorem A.4.3 (cf. Jacod and Shiryaev [23, Theorem III.5.32]), for all $t \geq 0$, it allows us to write the density process of P_θ with respect to $P_{\theta'}$ as

$$\frac{dP_\theta^t}{dP_{\theta'}^t} = \frac{dP_\theta^0}{dP_{\theta'}^0} \exp\left\{ U(t) - \frac{\sigma^2}{2} \int_0^t \beta^2(s) ds \right\} \prod_{s<t} (1 + \Delta U(s))\, e^{-\Delta U(s)},$$

which proves the proposition.

■

Proof of the Proposition 2.2.4

In order to apply the Proposition A.5.4 to determine a version of the Hellinger process associated with the measures P_θ and $P_{\theta'}$, the following notation will be needed,

1. $\varphi_\alpha : \mathbb{R}_+^2 \to \mathbb{R}_+ \ (\alpha \in (0,1))$

$$\varphi_\alpha(u,v) = \alpha u + (1-\alpha)v - u^\alpha v^{1-\alpha};$$

2. $C = c \cdot A$ and $C' = c' \cdot A$, for $A = \{A(t); t \geq 0\}$;

3. λ is a predictable random measure on $\mathbb{R}_+ \times \mathbb{R}^d$ such that

$$\left(|x|^2 \wedge 1\right) * \lambda_t < \infty, \quad \forall t < \infty,$$

$$\nu \ll \lambda \ \text{ and } \ \nu' \ll \lambda;$$

4. $\mathbf{U} = \{U(t); t \geq 0\}$ is a non-negative predictable function such that

$$\nu = U \cdot \lambda;$$

5. Σ is a predictable random set such that

$$\Sigma = \{(\omega,t); |h(x)(U-U')| * \lambda_t(\omega) < \infty\},$$

   where h is the truncation function;

6. $a_t(\omega) = \nu\left(\omega; \{t\} \times \mathbb{R}^d\right) = \int U(t,x)\lambda(\{t\} \times dx) \leq 1$.

Under this notations, taking $\nu = \nu' = \lambda$, we have $c = c' = \sigma^2$, $U \equiv 1$, $U' \equiv 1$, $\Sigma = \Omega \times \mathbb{R}_+$, $\varphi_\alpha(U,U') = \alpha + (1-\alpha) - 1 \equiv 0$ and $\varphi_\alpha(1-a(s), 1-a'(s)) = 0$.

Lemma 2.5.1. *Suppose that local uniqueness holds for the martingale problems* $s\left(\mathbf{X}|\, P_\theta^0; B, C, \nu\right)$ *and* $s\left(\mathbf{X}|\, P_{\theta'}^0; B', C, \nu\right)$. *Then, under Assumptions 2.2.1, a version of the Hellinger process* $\mathbf{H}(\alpha;\theta,\theta') = \{H(\alpha;\theta,\theta',t);\ t \geq 0\}$, $\alpha \in (0,1)$, *corresponding to the solution measures* P_θ *and* $P_{\theta'}$ *of the GLE (1.18) is given by*

$$H(\alpha;\theta,\theta',t) = \frac{\alpha(1-\alpha)}{2}\frac{1}{\sigma^2}\int_0^t \left(b(\theta,X_s) - b(\theta',X_s)\right)^2 ds,$$

provided that $\mathbf{H}(1/2;\theta,\theta')$ *does not jump to infinity.*

Proof. Denote by (B,C,ν) and (B',C',ν') the semimartingale characteristics of $\mathbf{X}$ under P_θ and $P_{\theta'}$ and let P_θ^0 and $P_{\theta'}^0$ be the initial distributions, respectively. By Proposition A.5.4 the problem of finding a measure to the GLE (1.18) is equivalent to solve the martingale problem $s\,(\mathbf{X}|\pi; B, C, \nu)$.

For each pair $\theta, \theta' \in \Theta$, define the predictable processes

$$\tilde{B}(t) = B(t) - B'(t) = \int_0^t \left(b(\theta,X_s) - b(\theta',X_s)\right) ds$$

and

$$\beta(t) = \beta(\theta,\theta',X_t) = \frac{b(\theta,X_t) - b(\theta',X_t)}{\sigma^2}.$$

Assuming that local uniqueness holds (see Definition A.4.2) for the martingale problems $s\left(\mathbf{X}|\, P_\theta^0; B, C, \nu\right)$ and $s\left(\mathbf{X}|\, P_{\theta'}^0; B', C, \nu\right)$ and $\mathbf{H}(1/2;\theta,\theta')$ does not jump to infinity, since $\nu = \nu' = \lambda$, by Equation (A.7) in Proposition A.5.4, we have that a version of the Hellinger process is given by

$$H(\alpha;\theta,\theta',t) = \frac{\alpha(1-\alpha)}{2}\frac{1}{\sigma^2}\int_0^t \left(b(\theta,X_s) - b(\theta',X_s)\right)^2 \mathbf{1}_\Sigma ds,$$

where $\Sigma = \Omega \times \mathbb{R}_+$.

■

Lemma 2.5.2. *Consider a family of measures* $\{P_\theta; \theta \in \Theta\}$ *induced by (1.18). Let* $\mathbf{L} = \{L(t); t \geq 0\}$ *be a Lévy process with characteristic triplet* (b, σ^2, μ)*. Under Assumption 2.2.3, for all pair* $(\theta, \theta') \in \Theta^2$*, we have that*

$$s\int_0^s \left(\int_0^u \rho_{\theta'}(u-v)dL(v)\right)^2 \mu_{\theta',s}(du) \in L^1\left([0,t]\times\Omega, ds \times dP_\theta\right).$$

Proof. We recall that by Lévy-Itô's decomposition, the Lévy process $\mathbf{L}$ can be written as a sum of a martingale $\{M(t); t \geq 0\}$ and an adapted process of finite variation $\mathbf{V} = \{V(t); t \geq 0\}$ such that

$$L(t) = \left(\sigma W(t) + \int_{|x|<1} x\widetilde{N}(t,dx)\right) + \left(bt + \int_{|x|\geq 1} xN(t,dx)\right) = M(t) + V(t),$$

where $\{W(t); t \geq 0\}$ is a Wiener process, $M(t) = \sigma W(t) + \int_{|x|<1} x\widetilde{N}(t,dx)$ and $V(t) = bt + \int_{|x|\geq 1} xN(t,dx)$.

Since $\rho_{\theta'}(\cdot)$ is predictable, for each $u \geq 0$, by [2, Section 4.3], $\int_0^u \rho_{\theta'}(u-v)dL(v)$ can be defined as

$$\int_0^u \rho_{\theta'}(u-v)dL(v) = \int_0^u \rho_{\theta'}(u-v)dM(v) + \int_0^u \rho_{\theta'}(u-v)dV(v).$$

It follows from (A.12) in Burkholder-Davis-Gundy's inequalities that

$$\begin{aligned}
E_\theta\left[\left(\int_0^u \rho_{\theta'}(u-v)dM(v)\right)^2\right] &\leq C_2 E_\theta\left[\int_0^u \rho_{\theta'}^2(u-v)d[M](v)\right] \\
&\leq C_2 \sup_{y\in[0,t]} \rho_{\theta'}^2(y) E_\theta\left[\int_0^u d[M](v)\right] \\
&= \bar{C}_2 E_\theta\left[[M](u)\right] \\
&= \bar{C}_2\left(E_\theta\left[[\sigma W](u) + \left[\int_{|x|<1} x\widetilde{N}(\cdot,dx)\right](u)\right]\right) \\
&= \bar{C}_2\left(\sigma^2 u + E_\theta\left[\sum_{v\leq u}(\Delta L(v))^2 \mathbf{1}_{[|\Delta L(v)|\leq 1]}\right]\right) \\
&= \bar{C}_2\left(\sigma^2 u + u\int_0^1 x^2\mu(dx)\right) \\
&\leq Cu, \qquad (2.19)
\end{aligned}$$

where in the second inequality we have used $\rho_{\theta'}(\cdot)$ is a continuous function in $[0,t]$ and $u \leq t$. Thus $\bar{C}_2 = C_2 \sup_{y\in[0,t]} \rho_{\theta_0}^2(y)$ is a positive (finite) constant. And in the last inequality we have used the fact that μ is a Lévy measure, so the integral of $x^2 \wedge 1$ with respect to this measure is finite. Here C is a positive constant.

Under Assumption 2.2.3, for all $t \geq 0$, the inequality (2.19) implies that

$$s \int_0^s E_\theta \left[\left(\int_0^u \rho_{\theta'}(u-v)dM(v) \right)^2 \right] \mu_{\theta',s}(du) \quad \leq \quad Cs \int_0^s u\mu_{\theta',s}(du) \in L^1\left([0,t], ds\right).$$

It follows from Fubini's Theorem that

$$s \int_0^s \left(\int_0^u \rho_{\theta'}(u-v)dM(v) \right)^2 \mu_{\theta',s}(du) \in L^1\left([0,t] \times \Omega, ds \times dP_\theta\right).$$

Finally, since $\mathbf{V}$ is of finite variation, we have that $V(t)$ is locally bounded. Thus, $V(u) \leq C_s \leq C_t$ for positive constants $C_s \leq C_t$, where $u \in [0,s] \subset [0,t]$. Again by the continuity of $\rho_{\theta_0}(\cdot)$ and Assumption 2.2.3, we obtain that

$$\begin{aligned}
E_\theta \left[\int_0^t s \int_0^s \left(\int_0^u \rho_{\theta'}(u-v)dV(v) \right)^2 \mu_{\theta',s}(du)ds \right] &\leq \sup_{[0,t]} \rho_{\theta'}^2 E_\theta \left[\int_0^t s \int_0^s \left(\int_0^u dV(v) \right)^2 \mu_{\theta',s}(du)ds \right] \\
&= \sup_{[0,t]} \rho_{\theta'}^2 E_\theta \left[\int_0^t s \int_0^s V^2(u) d\mu_{\theta',s}(u)ds \right] \\
&\leq \sup_{[0,t]} \rho_{\theta'}^2 C_t^2 E_\theta \left[\int_0^t s\mu_{\theta',s}([0,s])ds \right] \\
&\leq \sup_{[0,t]} \rho_{\theta'}^2 C_t^2 \int_0^t s\mu_{\theta',s}([0,s])ds < \infty,
\end{aligned}$$

for all fixed $t \geq 0$, which concludes the proof of the lemma.

■

Remark 2.5.3. *1. If $\mathbf{L} = \mathbf{W} = \{W(t); t \geq 0\}$ is a Wiener process, by [2, Lemma 4.3.11], we have that*

$$\int_0^t f(s)dL(s) \sim N\left(0, \int_0^t |f(s)|^2 ds\right), \quad t > 0.$$

Thus, for all $u \geq 0$,

$$E_\theta \left[\left(\int_0^u \rho_{\theta'}(u-v)dL(v) \right)^2 \right] = \int_0^u |\rho_{\theta'}(u-v)|^2 dv < \infty$$

and we can show the statement of the Lemma 2.5.2 more easily.

2. In general, $\int_0^t f(s)dL(s)$ cannot be defined as a Stieltjes integral. Actually only the continuous martingales that are of finite variation are constants (cf. Applebaum [2, p.138]).

For all $\theta \in \Theta$ and $t > 0$, we have that $\mu_{\theta,t}$ is a finite measure. This property assures us that we can apply Jensen's Inequality to obtain the Proposition 2.2.4.

Proof of the Proposition 2.2.4. By Theorem A.5.2, taking $T = t$, we have to show that $P_\theta-$almost sure $H(1/2; \theta, \theta', t) < \infty$ and $H(0; \theta, \theta', t) < \infty$. Lemma 2.5.1 gives us that $H(0; \theta, \theta', t) = 0$ and

$$H(1/2; \theta, \theta', t) = \frac{\alpha(1-\alpha)}{2} \frac{1}{\sigma^2} \int_0^t \beta^2(s)ds,$$

where $\beta(s) = \beta(\theta, \theta', X_s) = b(\theta, X_s) - b(\theta', X_s)$. It is enough to show that $\beta \in L^2\left([0,t] \times \Omega, ds \times dP_\theta\right)$. By the definition of the function $\beta(\cdot, \cdot)$, it suffices to prove that for every fixed pair $\theta, \theta' \in \Theta$, we have $b(s) = b(\theta', X_s) \in L^2\left([0,t] \times \Omega, ds \times dP_\theta\right)$.

By (2.3), for each $u \geq 0$ and each $\theta' \in \Theta$,

$$X(u) = X_{\theta'}(u) = X_0\rho_{\theta'}(u) + \int_0^u \rho_{\theta'}(u-v)dL(v).$$

it follows that we can rewrite the function $b(\theta', X_s)$ as

$$\begin{aligned} b(\theta', X_s) &= -\int_0^s X(u)\mu_{\theta',s}(du) \\ &= -X_0\int_0^s \rho_{\theta'}(u)\mu_{\theta',s}(du) - \int_0^s\int_0^u \rho_{\theta'}(u-v)dL(v)\mu_{\theta',s}(du) \\ &= X_0\rho'_{\theta'}(s) - \int_0^s\int_0^u \rho_{\theta'}(u-v)dL(v)\mu_{\theta',s}(du), \end{aligned}$$

where in the last equality we used (1.13). Thus,

$$\begin{aligned} b^2(\theta', X_s) &= \left(X_0\rho'_{\theta'}(s) - \int_0^s\int_0^u \rho_{\theta'}(u-v)dL(v)\mu_{\theta',s}(du)\right)^2 \\ &\leq 2X_0^2\left(\rho'_{\theta'}(s)\right)^2 + 2\left(\int_0^s\int_0^u \rho_{\theta'}(u-v)dL(v)\mu_{\theta',s}(du)\right)^2 \\ &= 2X_0^2\left(\rho'_{\theta'}(s)\right)^2 + 2s^2\left(\frac{1}{s}\int_0^s\int_0^u \rho_{\theta'}(u-v)dL(v)\mu_{\theta',s}(du)\right)^2 \\ &\leq 2X_0^2\left(\rho'_{\theta'}(s)\right)^2 + 2s^2\frac{1}{s}\int_0^s\left(\int_0^u \rho_{\theta'}(u-v)dL(v)\right)^2\mu_{\theta',s}(du) \\ &= 2X_0^2\left(\rho'_{\theta'}(s)\right)^2 + 2s\int_0^s\left(\int_0^u \rho_{\theta'}(u-v)dL(v)\right)^2\mu_{\theta',s}(du), \qquad (2.20) \end{aligned}$$

where the second inequality is obtained applying Jensen's Inequality for finite measures.

Both terms in the right side of (2.20) are functions in $L^1([0,t]\times\Omega, ds\times dP_\theta)$. Indeed, the second term in (2.20) is guaranteed by the Lemma 2.5.2. For the first term, note that

$$E_\theta\left[X_0^2\left(\rho'_{\theta'}(s)\right)^2\right] = \left(\rho'_{\theta'}(s)\right)^2 E_\theta\left[X_0^2\right].$$

Thus,

$$\int_0^t E_\theta\left[X_0^2\left(\rho'_{\theta'}(s)\right)^2\right]ds = E_\theta\left[X_0^2\right]\int_0^t\left(\rho'_{\theta'}(s)\right)^2 ds < \infty$$

and applying Fubini's Theorem, we conclude that

$$E_\theta\left[\int_0^t X_0^2\left(\rho'_{\theta'}(s)\right)^2 ds\right] < \infty,$$

which proves that $X_0^2\left(\rho'_{\theta'}\right)^2 \in L^1([0,t]\times\Omega, ds\times dP_\theta)$ and concludes the proof of the proposition.

■

Proof of the Theorem 2.3.4

We have divided the proof of the Theorem 2.3.4 in a sequence of lemmas. This prove is based on Gloter, Loukianova and Mai [13].

Lemma 2.5.4. *Let* $\mathbf{M} = \{M(t); t \geq 0\}$ *be the continuous local* $P_{\theta_0}-$*martingale defined by*

$$M(\theta, t) := \frac{1}{\sigma}\int_0^t b(\theta, X_s) - b(\theta_0, X_s) dW(s),$$

where $\mathbf{W} = \{W(t); t \geq 0\}$ *is a* P_{θ_0}*-Wiener process. Define the function*

$$\mathcal{L}(\theta; X_t) := M(\theta, t) - \frac{1}{2}[M(\theta)](t), \tag{2.21}$$

in which $[M(\theta)](t)$ *denotes the quadratic variation of* $\mathbf{M}$*. Then*

$$\hat{\theta}(t) \in \arg\max_{\theta \in \Theta_0} \mathcal{L}(\theta, X_t)\cdot$$

Proof. Under P_{θ_0}, the continuous martingale part of X is given by

$$X^c(t) = \sigma W(t) + \int_0^t b(\theta_0, X_s) ds\cdot \tag{2.22}$$

The quadratic variation of $\mathbf{M}$ is

$$[M(\theta)](t) = \frac{1}{\sigma^2}\int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right)^2 ds\cdot$$

Being θ_0 the true value of the parameter corresponding to the observed path of $\mathbf{X}$, by (2.8) and (2.22), we have

$$\begin{aligned} l(\theta, X_t) &= \frac{1}{\sigma}\int_0^t b(\theta, X_s) dW(s) + \frac{1}{\sigma^2}\int_0^t b(\theta, X_s) b(\theta_0, X_s) ds - \frac{1}{2\sigma^2}\int_0^t b^2(\theta, X_s) ds \\ &= \frac{1}{\sigma}\int_0^t b(\theta, X_s) - b(\theta_0, X_s) dW(s) - \frac{1}{2\sigma^2}\int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right)^2 ds \\ &+ \frac{1}{\sigma}\int_0^t b(\theta_0, X_s) dW(s) + \frac{1}{2\sigma^2}\int_0^t b^2(\theta_0, X_s) ds \\ &= \mathcal{L}(\theta; X_t) + \frac{1}{\sigma}\int_0^t b(\theta_0, X_s) dW(s) + \frac{1}{\sigma^2}\int_0^t b^2(\theta_0, X_s) ds\cdot \end{aligned}$$

Thus, the difference $l(\theta, X_t) - \mathcal{L}(\theta, X_t)$ does not depend on θ, which implies $\hat{\theta}(t) \in \arg\max_{\theta \in \Theta_0} \mathcal{L}(\theta, X_t)$.

∎

Lemma 2.5.5. *Suppose that Assumption 2.3.1 holds. Then,* $P_{\theta_0}-$*a.s. on any compact* $K \subset \Theta_0$ *not containing* θ_0*,*

$$\lim_{t \to \infty} \sup_{\theta \in K} \left| \frac{1}{\varphi(t)} \mathcal{L}(\theta, X_t) - \mathcal{L}(\theta) \right| = 0 \tag{2.23}$$

where, by the limit in (2.11), we define $\mathcal{L}(\theta) := -\frac{1}{2}\tilde{\Sigma}^2(\theta)\cdot$

Proof. Let us first prove that the family

$$\left\{ \frac{1}{\varphi(t)}[M(\theta)](t);\ (\theta, t) \in \Theta_0 \times \mathbb{R}_+ \right\}$$

is bounded and equicontinuous on θ, $P_{\theta_0}-$a.s. Indeed, by Assumption 2.3.1, $P_{\theta_0}-$a.s.

$$\lim_{t \to \infty} -\frac{1}{2\varphi(t)}[M(\theta)](t) = -\frac{1}{2}\tilde{\Sigma}^2(\theta) =: \mathcal{L}(\theta)\cdot \tag{2.24}$$

Furthermore, since Θ_0 is compact, there exists a constant $C > 0$ for which all $\theta, \theta' \in \Theta_0$ belong to a ball of radius C and center θ_0. Under Assumption 2.3.1, for every $\epsilon > 0$ take $\delta = \delta(\epsilon) = \frac{\epsilon\sigma^2}{2C^\kappa \Sigma_K^2} > 0$ such that, for all $\theta, \theta' \in \Theta_0$, if $|\theta - \theta'|^\kappa < \delta$, then

$$\left|[M(\theta)](t) - [M(\theta')](t)\right| = \left|\int_0^t \left[(b(\theta, X_s) - b(\theta_0, X_s))^2 - \left(b(\theta', X_s) - b(\theta_0, X_s)\right)^2\right] ds\right|,$$

which implies

$$\begin{aligned}
\frac{1}{\varphi(t)}\left|[M(\theta)](t) - [M(\theta')](t)\right| &\leq \frac{1}{\sigma^2\varphi(t)}\int_0^t \left|b(\theta, X_s) - b(\theta', X_s)\right| \left|b(\theta, X_s) - b(\theta_0, X_s)\right| ds \\
&+ \frac{1}{\sigma^2\varphi(t)}\int_0^t \left|b(\theta, X_s) - b(\theta', X_s)\right| \left|b(\theta', X_s) - b(\theta_0, X_s)\right| ds \\
&\leq |\theta - \theta'|^\kappa \left(|\theta - \theta_0|^\kappa + \left|\theta' - \theta_0\right|^\kappa\right) \frac{1}{\sigma^2\varphi(t)}\int_0^t K^2(X_s) ds \\
&\leq |\theta - \theta'|^\kappa \frac{2C^\kappa}{\sigma^2\varphi(t)}\int_0^t K^2(X_s) ds \\
&\leq |\theta - \theta'|^\kappa \frac{2C^\kappa}{\sigma^2}\Sigma_K^2 < \epsilon, \quad \text{for all } \ t > 0, \ \ P_{\theta_0} - a.s.
\end{aligned}$$

where Σ_K^2 is given in Assumption 2.3.1, and this proves our claim. It follows from Arzelá-Ascoli Theorem that $P_{\theta_0}-$a.s.

$$\lim_{t\to\infty} \sup_{\theta\in\Theta_0} \left| -\frac{1}{2\varphi(t)}[M(\theta)](t) - \mathcal{L}(\theta)\right| = 0\cdot \tag{2.25}$$

Moreover,

$$[M(\theta) - M(\theta')](t) = \frac{1}{\sigma^2}\int_0^t \left(b(\theta, X_s) - b(\theta', X_s)\right)^2 ds \leq V(t)|\theta - \theta'|^{2\kappa},$$

where $V(t) = \frac{1}{\sigma^2}\int_0^t K^2(X_s)ds \uparrow \infty$ $P_{\theta_0}-$a.s. Thus, the conditions of Theorem A.7.6 are satisfied. Then, the family $\left\{\frac{M(\theta,t)}{[M(\theta)](t)};\ (\theta, t) \in \Theta_0 \times \mathbb{R}_+\right\}$ satisfies the uniform law of large numbers on any compact $K \subset \Theta_0$ not containing θ_0, that is

$$\lim_{t\to\infty} \sup_{\theta\in K} \left|\frac{M(\theta, t)}{[M(\theta)](t)}\right| = 0 \quad P_{\theta_0} - a.s. \tag{2.26}$$

Observe that by (2.25) and (2.26) we obtain

$$\lim_{t\to\infty} \sup_{\theta\in K} \left|\frac{M(\theta, t)}{\varphi(t)}\right| = 0 \quad P_{\theta_0} - a.s. \tag{2.27}$$

Therefore, by (2.21), (2.25) and (2.27), the limit (2.23) holds true.

■

Before we establish the next lemma, we will recall a version of the Taylor's theorem for multivariate functions.

Let $f : \mathbb{R}^N \to \mathbb{R}$ be a $(k + 1)-$times continuously differentiable at a point $a \in \mathbb{R}^N$. Then

$$f(x) = f(a) + \sum_{\alpha\in\mathbb{N}^N; 1\leq|\alpha|\leq k} \frac{1}{\alpha!} D^\alpha f(a)(x-a)^\alpha + \sum_{\alpha\in\mathbb{N}^N; |\alpha|=k+1} (x-a)^\alpha \int_0^1 (1-s)^k D^\alpha f\left(a + s(x-a)\right) ds,$$

where for $x \in \mathbb{R}^N$ and $\alpha^\top = (\alpha_1, \cdots, \alpha_N) \in \mathbb{N}^N$, we denote $\alpha! = \prod_{j=1}^N \alpha_j!$, $x^\alpha = \prod_{j=1}^N x_j^{\alpha_j}$, $|\alpha| = \sum_{j=1}^N \alpha_j$ and

$$D^\alpha f(x) = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_N^{\alpha_N}}(x)\cdot$$

Lemma 2.5.6. *Under Assumption 2.3.2 we have the decomposition*

$$\frac{1}{\varphi(t)} \int_0^1 \partial_\theta^2 \mathcal{L}\left(\theta_0 + s\left(\hat{\theta}(t) - \theta_0\right), X_t\right) ds \times \sqrt{\varphi(t)}\left(\hat{\theta}(t) - \theta_0\right) = -\frac{1}{\sqrt{\varphi(t)}} \nabla_\theta \mathcal{L}\left(\theta_0, X_t\right)\cdot \quad (2.28)$$

Furthermore,

$$Law\left(\frac{1}{\sqrt{\varphi(t)}} \nabla_\theta \mathcal{L}\left(\theta_0, X_t\right) \middle| P_{\theta_0}\right) \to N(0, I(\theta_0)), \ \ as \ \ t \to \infty, \quad (2.29)$$

where $I(\theta_0)$ is defined in (2.12).

Proof. It follows from Assumption 2.3.2, interchanging the derivation with respect θ and the integrals in (2.21), that

$$\nabla_\theta \mathcal{L}\left(\theta, X_t\right) = \left(\partial_{\theta_1} \mathcal{L}\left(\theta, X_t\right), \cdots, \partial_{\theta_N} \mathcal{L}\left(\theta, X_t\right)\right)^\top,$$

where, for each $j = 1, \cdots, N$,

$$\partial_{\theta_j} \mathcal{L}(\theta, X_t) = -\frac{1}{\sigma^2} \int_0^t \left(b(\theta, X_s) - b(\theta_0, X_s)\right) \partial_{\theta_j} b(\theta, X_s) ds + \frac{1}{\sigma} \int_0^t \partial_{\theta_j} b(\theta, X_s) dW(s)\cdot \quad (2.30)$$

By doing a Taylor expansion of $\nabla_\theta \mathcal{L}(\cdot, X_t)$ around θ_0 and applying it in $\hat{\theta}(t)$ we obtain (2.28). Actually, for each $j = 1, \cdots, N$ and each $\theta \in \text{int}\, \Theta_0$,

$$\partial_{\theta_j} \mathcal{L}(\theta, X_t) = \partial_{\theta_j} \mathcal{L}(\theta_0, X_t) + \sum_{i=1}^N (\theta_i - \theta_{0,i}) \int_0^1 \frac{\partial^2}{\partial_{\theta_j} \partial_{\theta_i}} \mathcal{L}\left(\theta_0 + s(\theta - \theta_0)\right) ds,$$

which implies

$$\nabla_\theta \mathcal{L}(\theta, X_t) = \nabla_\theta \mathcal{L}(\theta_0, X_t) + \int_0^1 \partial^2 \mathcal{L}(\theta_0 + s(\theta - \theta_0), X_t) ds \times (\theta - \theta_0)\cdot \quad (2.31)$$

Since $\hat{\theta}(t) \in \arg\max_{\theta \in \Theta_0} \mathcal{L}(\theta, X_t)$, we have $\nabla_\theta \mathcal{L}\left(\hat{\theta}(t), X_t\right) = 0$ and, therefore, applying (2.31) in $\hat{\theta}(t)$, (2.28) follows.

Furthermore, for each $j = 1, \cdots, N$,

$$\partial_{\theta_j} \mathcal{L}(\theta_0, X_t) = \frac{1}{\sigma} \int_0^t \partial_{\theta_j} b(\theta_0, X_s) dW(s),$$

that yields

$$\frac{1}{\sqrt{\varphi(t)}} \nabla_\theta \mathcal{L}\left(\theta_0, X_t\right) = \frac{1}{\sigma\sqrt{\varphi(t)}} \int_0^t \nabla_\theta b(\theta_0, X_s) dW(s), \quad (2.32)$$

therefore, (2.29) follows from Assumption 2.3.2 and the CLT for N-dimensional martingales A.7.4 taking $K(t) = \varphi(t)^{-1/2} I_N$ and $M(t) = \nabla_\theta \mathcal{L}\left(\theta_0, X_t\right)\cdot$

■

Proof of Theorem 2.3.4. We divided the proof in tree steps.

Step 1. Strong consistency of $\hat{\theta}(t)$.

By Lemma 2.5.4 we know that the MLE $\hat{\theta}(t)$ of θ_0 also maximizes the function $\mathcal{L}(\theta, X_t)$. Furthermore, Lemma 2.5.5 proves the almost sure uniform convergence of $\varphi(t)^{-1}\mathcal{L}(\theta, X_t)$ to $\mathcal{L}(\theta)$. By the Wald's Method for proving consistence of estimators (cf. Vaart [65, Theorem 5.7] for a simple version of this method), it suffices to show that for all $\varepsilon > 0$

$$\lim_{t\to\infty} \sup_{\theta;|\theta-\theta_0|\geq\varepsilon} \varphi(t)^{-1}\mathcal{L}(\theta, X_t) < \mathcal{L}(\theta_0)\cdot \tag{2.33}$$

Indeed, by (2.24) we have that $\mathcal{L}(\theta) \leq 0$ and $\mathcal{L}(\theta) = 0$ if and only if $\theta = \theta_0$. Thus,

$$\sup_{\theta;|\theta-\theta_0|\geq\varepsilon} \mathcal{L}(\theta) < \mathcal{L}(\theta_0)\cdot \tag{2.34}$$

Then, (2.33) follows from (2.23) and (2.34). Hence, for P_{θ_0} almost all $\omega \in \Omega$ fixed and for all $t > t(\omega)$ large enough we have $\sup_{\theta;|\theta-\theta_0|\geq\varepsilon} \varphi(t)^{-1}\mathcal{L}(\theta, X_t) < \mathcal{L}(\theta_0, X_t)$, which implies $\left|\hat{\theta}(t) - \theta_0\right| < \varepsilon$ for $t > t(\omega)$.

Step 2. Asymptotic normality.

By (2.30), for all $(i,j) \in \{1, \cdots, N\}^2$

$$\begin{aligned}\partial^2_{\theta_i\theta_j}\mathcal{L}(\theta, X_t) &= -\frac{1}{\sigma^2}\int_0^t (b(\theta, X_s) - b(\theta_0, X_s))\, \partial^2_{\theta_i\theta_j} b(\theta, X_s) ds \\ &- \frac{1}{\sigma^2}\int_0^t \partial_{\theta_i} b(\theta, X_s)\partial_{\theta_j} b(\theta, X_s) ds + \frac{1}{\sigma}\int_0^t \partial^2_{\theta_i\theta_j} b(\theta, X_s) dW(s)\cdot \end{aligned} \tag{2.35}$$

By the same arguments as in the Lemma 2.5.5's proof, we can show that under Assumption 2.3.2 the families

$$\left\{\frac{1}{\varphi(t)}\int_0^t (b(\theta, X_s) - b(\theta_0, X_s))\, \partial^2_{\theta_i\theta_j} b(\theta, X_s) ds;\ (t,\theta) \in \mathbb{R}_+ \times \Theta_0\right\}$$

and $\left\{\frac{1}{\varphi(t)}\int_0^t \partial_{\theta_i} b(\theta, X_s)\partial_{\theta_j} b(\theta, X_s) ds;\ (t,\theta) \in \mathbb{R}_+ \times \Theta_0\right\}$ are pointwise bounded and equicontinuous on θ $P_{\theta_0}-$a.s. Then they converge uniformly. Moreover, the uniform law of large numbers Loukianova and Loukianov [40, Theorem 2] yields

$$\sup_{\theta\in\Theta_0}\left|\frac{1}{\varphi(t)}\partial^2_{\theta_i\theta_j}\mathcal{L}(\theta, X_t) - (\xi_{ij}(\theta) - I_{ij}(\theta))\right| \to 0 \quad P_{\theta_0} - a.s.,$$

which, together with $\xi(\theta_0) = 0$ from (2.13) and the consistence of $\hat{\theta}(t)$, implies

$$\sup_{s\in[0,1]}\left|\frac{1}{\varphi(t)}\partial^2_{\theta}\mathcal{L}\left(\theta_0 + s\left(\hat{\theta}(t) - \theta_0\right), X_t\right) + I(\theta_0)\right| \to 0$$

and

$$\frac{1}{\varphi(t)}\int_0^1 \partial^2_{\theta}\mathcal{L}\left(\theta_0 + s\left(\hat{\theta}(t) - \theta_0\right), X_t\right) ds \to -I(\theta_0), \tag{2.36}$$

$P_{\theta_0}-$a.s. as $t \to \infty\cdot$ Therefore, the asymptotic normality of the estimator follows from (2.36), Lemma 2.5.6 and Slutsky's lemma.

Step 3. LAN property.

Due to (2.29), for all $\theta_0 \in \Theta_0$ and $h \in \mathbb{R}^N$ it is sufficient to show that, under P_{θ_0},

$$\log \frac{dP^t_{\theta+h\varphi(t)^{-1/2}}}{dP^t_\theta} = \frac{h^\top}{\sqrt{\varphi(t)}} \nabla_\theta \mathcal{L}(\theta_0, X_t) - \frac{h^\top}{2\sigma^2\varphi(t)} \int_0^t \nabla_\theta b(\theta_0, X_s) \nabla_\theta b(\theta_0, X_s)^\top ds h + o_{P_{\theta_0}}(1),$$

where $\nabla_\theta \mathcal{L}(\theta_0, X_t)$ is given in (2.32). Actually, (2.4), (2.22) and (2.32) gives us

$$\begin{aligned} \log \frac{dP^t_{\theta+h\varphi(t)^{-1/2}}}{dP^t_\theta} &= \frac{h^\top}{\sqrt{\varphi(t)}} \nabla_\theta \mathcal{L}(\theta_0, X_t) - \frac{1}{2\sigma^2\varphi(t)} \int_0^t h^\top \nabla_\theta b(\theta_0, X_s) \nabla_\theta b(\theta_0, X_s)^\top h ds \\ &+ R_1(t) + R_2(t), \end{aligned}$$

where

$$R_1(t) = \frac{1}{\sigma} \int_0^t b\left(\theta_0 + h\varphi(t)^{-1/2}, X_s\right) - b(\theta_0, X_s) dW(s) - \frac{1}{\sigma\sqrt{\varphi(t)}} \int_0^t h^\top \nabla_\theta b(\theta_0, X_s) dW(s)$$

and

$$\begin{aligned} R_2(t) &= \frac{1}{2\sigma^2\varphi(t)} \int_0^t h^\top \nabla_\theta b(\theta_0, X_s) \nabla_\theta b(\theta_0, X_s)^\top h ds \\ &+ \frac{1}{\sigma} \int_0^t \left(b(\theta_0 + h\varphi(t)^{-1/2}, X_s) - b(\theta_0, X_s) \right) b(\theta_0, X_s) ds \\ &- \frac{1}{2\sigma^2} \int_0^t \left(b(\theta_0 + h\varphi(t)^{-1/2}, X_s) - b(\theta_0, X_s) \right)^2 ds. \end{aligned}$$

Observe that, for $\theta_0, \theta \in \Theta$ and $s \geq 0$,

$$b\left(\theta_0 + h\varphi(t)^{-1/2}, X_s\right) - b(\theta, X_s) = \frac{h^\top}{\sqrt{\varphi(t)}} \int_0^1 \nabla_\theta b\left(\theta_0 + h\varphi(t)^{-1/2} u, X_s\right) du \tag{2.37}$$

implies, for a constant $C > 0$,

$$E_{\theta_0}\left[|R_1(t)|^2\right] \leq \frac{|h|^{2+2\kappa_1} C}{\varphi(t)^{\kappa_1}} E_{\theta_0}\left[\frac{1}{\varphi(t)} \int_0^t K_1^2(X_s) ds\right]$$

which converges to zero. Then, the Markov inequality implies that $R_1(t) = o_{P_{\theta_0}}(1)$.

$$\begin{aligned} |R_2(t)| &= \left| \frac{1}{2\sigma^2\varphi(t)} \int_0^t h^\top \nabla_\theta b(\theta_0, X_s) \nabla_\theta b(\theta_0, X_s)^\top h ds \right. \\ &- \left. \frac{1}{2\sigma^2\varphi(t)} \int_0^t \int_0^1 \int_0^1 h^\top \nabla_\theta b(\theta_0 + h\varphi(t)^{-1/2} u, X_s) \nabla_\theta b(\theta_0 + h\varphi(t)^{-1/2} u', X_s)^\top h du du' ds \right| \\ &\leq \frac{|h|^2}{2\sigma^2} \int_0^1 \int_0^1 \frac{1}{\varphi(t)} \int_0^t \left| \nabla_\theta b(\theta_0, X_s) \nabla_\theta b(\theta_0, X_s)^\top \right. \\ &- \left. \nabla_\theta b(\theta_0 + h\varphi(t)^{-1/2} u, X_s) \nabla_\theta b(\theta_0 + h\varphi(t)^{-1/2} u', X_s)^\top \right| ds du du'. \end{aligned}$$

Therefore, $R_2(t) = o_{P_{\theta_0}}(1)$ and the LAN property holds true.

■

Chapter 3

A Three-Parameter GOU-FE Process and a Modified FMLE

3.1 Introduction

In Chapter 2, we studied the FMLE

$$\hat{\theta}_T^{FMLE} := \arg\max_{\theta\in\Theta_0}\left\{\frac{1}{\sigma^2}\sum_{i=1}^{n} b(\theta, X_{t_i})\Delta_i X \mathbf{1}_{[|\Delta_i X|\leq v_n^i]} - \frac{1}{2\sigma^2}\sum_{i=1}^{n} b^2(\theta, X_{t_i})\Delta_i\right\} \tag{3.1}$$

for the drift parameter of a GLE for which the solution is the GOU-FE process. Our simulation results for the process

$$X(t) = X_0(0)\left((1-\theta_2)e^{-\theta_1 t} + \theta_2\cos(\theta_2 t)\right) + \int_0^t \left((1-\theta_2)e^{-\theta_1(t-s)} + \theta_2\cos(\theta_2(t-s))\right) dL(s)$$

showed that despite the good results obtained when the perturbation of the OU process is small, i.e., $\theta_2 \downarrow 0$, issues such as computational time and improved estimation in some regions of parametric space Θ_0 require a little more care, so that this estimator can be applied to model real data via GLE.

In this chapter, we propose a new form of discretization for the MLE $\hat{\theta}(t) := \arg\max_{\theta\in\Theta_0} l(\theta, X_t)$. By considering all the past information of the process $\mathbf{X}$ until the time t_k, that is,

$$X_{t_k} = \{X(t_1), \cdots, X(t_k)\},$$

one can consider a discretization of the likelihood function using the jump filtering technique to approximate the increments of the continuous martingale part X^c. This will lead us to introduce a modified FMLE (mFMLE). In Section 3.3, we propose the new estimator mFMLE that make use of the information of the past simulated noise and the real expected noise,

$$\hat{\theta}_T^{mFMLE} = \arg\min_{\theta\in\Theta_0} \left| l(\theta, X_{t_n}) - l(\theta_0, X_{t_n})\right|,$$

where θ_0 is the true unknown parameter value and $\{t_1, \cdots, t_n\}$ is a convenient partition of the time interval. Simulation results for mFMLE and its performance is compared to that of FMLE

In Section 3.2, we generalize the GOU-FE process, in the sense that we give more freedom for the exponential and cosine functions in $\rho_\theta(t) = (1-\theta_2)e^{-\theta_1 t} + \theta_2\cos(\theta_1 t)$. Then, we study the corresponding Kannan's solution

$$X(t) = X_0(0)\rho_\theta(t) + \int_0^t \rho_\theta(t-s)dL(s), \ \ t > 0, \tag{3.2}$$

with

$$\rho_\theta(t) = (1-\theta_3)e^{-\theta_1 t} + \theta_3\cos(\theta_2 t),$$

where $\theta^\top = (\theta_1, \theta_2, \theta_3) \in \Theta \subset \mathbb{R}^2 \times [0,1]$. As in Alcântara [1], our Theorem 3.2.1 shows that this process is indeed a solution of the GLE with the associated family of signed measures $\{\mu_{\theta,t}; \theta \in \Theta,\ t \geq 0\}$ satisfying the Volterra integro-diferential

$$\begin{cases} \rho'_\theta(t) &= -\displaystyle\int_0^t \rho_\theta(s)\mu_{\theta,t}(ds), \\ \rho_\theta(0) &= 1. \end{cases} \tag{3.3}$$

Our Proposition 3.2.2 shows that a discretization of this new process has an order 3 autoregressive form, which is a consequence of the OU and cosine recurrences (see Stein, Lopes and Medino [62, Example 4.1 and Proposition 4.1]). Also, the Theorem 3.2.6 shows that if the "memory" function ρ satisfies the Assumption 3.2.4 then the corresponding OU process will have autoregressive representation of general order m. Thus the same discretization technique could be applied. Extension for the case when $\mathbf{L}$ is a symmetric $\alpha-$stable Lévy process with $1 < \alpha \leq 2$ is considered in Proposition 3.2.7.

The last section is dedicated to the proof of the results of this chapter. The main tool used to prove the Theorem 3.2.1 is the Laplace transform and its properties. The autoregressive properties are obtained, as mentioned above, by direct applications or by a careful review of some proofs from Stein, Lopes and Medino [62].

3.2 A New Class of Solution for the GLE

In this section, we extend the parameter space of the GOU-FE process from a subset of $\mathbb{R}^2$ to a subset of $\mathbb{R}^3$. The new process takes the form of the Kannan's solution (3.2) where

$$\rho_\theta(t) = (1-\theta_3)e^{-\theta_1 t} + \theta_3\cos(\theta_2 t), \tag{3.4}$$

$\theta^\top = (\theta_1, \theta_2, \theta_3) \in \Theta \subset \mathbb{R}_+ \times \mathbb{R} \times [0,1]$.

Taking $\theta_3 = 0$, $\theta_3 = 1$ or $\theta_2 = \theta_1$, we obtain, respectively, the processes Ornstein-Uhlenbeck, cosine (1.3.3) or the GOU-FE. As for these examples , the proposed GOU process with $\rho_\theta(\cdot)$ defined in (3.4) is a solution of a GLE's.

Theorem 3.2.1. *Let $\mathbf{L} = \{L(t); t \geq 0\}$ be a Lévy process with finite second moment or $\alpha-$stable with $1 < \alpha \leq 2$. Then, the process $\mathbf{X} = \{X(t); t \geq 0\}$ given by (3.2) with $\rho_\theta(\cdot)$ defined in (3.4) is a solution for the GLE*

$$dX(t) = \left(-\theta_1(1-\theta_3)X(t) - \int_0^t X(s)\Gamma_\theta(t-s)ds\right)dt + dL(t),\ \ t > 0,\ \ \textit{and}\ \ X(0) = X_0, \tag{3.5}$$

where

$$\Gamma_\theta(t) = \begin{cases} e^{-t\theta_1\theta_3/2}\left(\kappa_1\cos(\nu t) + \kappa_{2,1}\sin(\nu t)\right), & \nu > 0, \\ e^{-t\theta_1\theta_3/2}\left(\kappa_1\cosh(\nu t) + \kappa_{2,-1}\sinh(\nu t)\right), & \nu < 0, \\ e^{-t\theta_1\theta_3/2}\left(\kappa_1 + \kappa_{2,0}t\right), & \nu = 0, \end{cases}$$

for the constants

$$\begin{aligned}
\nu_0 &= \theta_1\theta_3/2,\\
\nu &= -\nu_0^2+\theta_2^2(1-\theta_3),\\
\kappa_1 &= \theta_2^2\theta_3-2\theta_1(1-\theta_3)\nu_0,\\
\kappa_{2,1} &= \frac{1}{\nu}\left(\theta_1\theta_2^2-\nu_0\left(\theta_2^2\theta_3-\theta_1(1-\theta_3)2\nu_0\right)-\theta_1(1-\theta_3)\left(\nu_0^2+\nu\right)\right),\\
\kappa_{2,-1} &= \frac{1}{\nu}\left(\theta_1\theta_2^2-\nu_0\left(\theta_2^2\theta_3-\theta_1(1-\theta_3)2\nu_0\right)-\theta_1(1-\theta_3)\left(\nu_0^2-\nu\right)\right),\\
\kappa_{2,0} &= \theta_1\theta_2^2-\nu_0\left(\theta_2^2\theta_3-\theta_1(1-\theta_3)2\nu_0\right)-\theta_1(1-\theta_3)\nu_0^2.
\end{aligned}$$

We point out that the proof of the Theorem 3.2.1 uses basically Laplace transform and finding the relations (3.3) between the memory function and $\rho_\theta(\cdot)$. The same result could be obtained via power series and Picard iterative method, as were done in Stein [61, Examples 2.2 and 2.3] for the cases $\rho_\theta(t)=\cos(\theta t)$ and $\rho_\theta(t)=e^{-\theta t^2}$.

Few notation will be needed. Denote the processes OU and cosine by $\{A(t);t\geq 0\}$ and $\{B(t);t\geq 0\}$, respectively. As mentioned in Examples 1.3.2 and 1.3.3, from Stein, Lopes and Medino [62, Example 4.1 and Proposition 4.1] the following recurrence properties hold,

$$A((k+1)h)=e^{-\theta_1 h}A(kh)+\varepsilon_{A,k+1}, \tag{3.6}$$

and

$$B((k+1)h)=2\cos(\theta_2 h)B(kh)-B((k-1)h)+\varepsilon_{B,k+1}, \tag{3.7}$$

where

$$\varepsilon_{A,k+1}=\int_{kh}^{(k+1)h}e^{-\theta_1((k+1)h-s)}dL(s) \tag{3.8}$$

and

$$\begin{aligned}
\varepsilon_{B,k+1} &= -\int_{(k-1)h}^{kh}\cos\left(\theta_2((k-1)h-s)\right)dL(s)\\
&+\int_{kh}^{(k+1)h}\left[2\cos(\theta_2 h)\cos(\theta_2(kh-s))-\cos\left(\theta_2((k-1)h-s)\right)\right]dL(s)\cdot
\end{aligned} \tag{3.9}$$

Similarly, we will show that a discrete time path of the solution process of (3.5) satisfies a recurrence form as given in the next Proposition. This result can provide us with a third simulation method: the GLE can be simulated by Euler-Maruyama and the integral process (3.2) can be simulated by Riemann sums.

Proposition 3.2.2. *Let* $\mathbf{X}=\{X(t);t\geq 0\}$ *be a process satisfying (3.2) with* $\rho_\theta(t)$ *defined in (3.4). For a fixed increment size* $h>0$ *the process* $\mathbf{X}$ *satisfies the recurrence relations*

$$X((k+1)h)=\phi_1 X(kh)+\phi_2 X((k-1)h)+\phi_3 X((k-2)h)+\varepsilon_{k,h} \tag{3.10}$$

where

$$\begin{aligned}
\phi_1 &= e^{-\theta_1 h}+2\cos(\theta_2 h),\\
\phi_2 &= -1-2\cos(\theta_2 h)e^{-\theta_1 h},\\
\phi_3 &= e^{-\theta_1 h},
\end{aligned}$$

and, for $\varepsilon_{A,k+1}$ and $\varepsilon_{B,k+1}$ given respectively by (3.8) and (3.9),

$$\begin{aligned}\varepsilon_{k,h} &= (1-\theta_3)\varepsilon_{A,k+1} - (1-\theta_3)2\cos(\theta_2 h)\varepsilon_{A,k} + (1-\theta_3)\varepsilon_{A,k-1} \\ &+ \theta_3\varepsilon_{B,k+1} - \theta_3 e^{-\theta_1 h}\varepsilon_{B,k}.\end{aligned}$$

Note that the errors $\varepsilon_{k,h}$ are not i.i.d., much less a white noise, hampering the use of classical estimation techniques for autoregressive processes. Alcântara [1] worked around this by decomposing the error $\varepsilon_{k,h}$ and noting that the process would have characteristics of an ARMA (*autoregressive moving average*) model.

We now want to obtain a general form of recurrence for a GOU process satisfying (3.2). Note that the recurrence of the processes OU and cosine were obtained by Stein, Lopes and Medino [62] using the recurrence of the function $\rho_\theta((k+1)h)$ in the exponential and cosine cases. However, Alcântara [1] obtained the recurrence of the GOU-FE through the previously established recurrences of OU and cosine processes. This approach, despite being natural for the construction of the GOU-FE, is more difficult to be generalized for a GOU process. Thus, we would like to obtain a recurrence form for the GOU (3.2) requiring conditions only in the discrete form of $\rho_\theta(\cdot)$.

The following example was implicitly used by Stein [61] and Stein, Lopes and Medino [62] in the proofs of the recurrence forms for the studied processes. It gives us an important idea of what type of property to require in order to get recurrence relations for the GOU process (3.2).

Example 3.2.3. *1. If $\rho_\theta(t) = e^{-\theta t}$, then*

$$\rho_\theta\left((k+1)h - s\right) = e^{-\theta h}\rho_\theta\left(kh - s\right);$$

2. if $\rho_\theta(t) = \cos(\theta t)$, then

$$\rho_\theta\left((k+1)h - s\right) = 2\cos(\theta h)\rho_\theta\left(kh - s\right) - \rho_\theta\left((k-1)h - s\right);$$

3. if $\rho_\theta(t) = e^{-\theta t^2}$, then

$$\rho_\theta\left((k+1)h - s\right) = e^{-\theta((2k+1)h^2 - 2hs)}\rho_\theta\left(kh - s\right).$$

Note that 3 in Example 3.2.3 gives us the idea of how to use a similar form of a recurrent function, that is, to accept that the recurrence constants can depend on k and h. Consider that $\rho_\theta(\cdot)$ satisfies the following condition.

Assumption 3.2.4. *For fixed values of $m, k \in \mathbb{N}$, $h, s \in \mathbb{R}_+$ and $\theta \in \Theta$, there exists a non-null constant $\alpha_j(s) = \alpha_j(\theta, (k+1)h - s)$ such that*

$$\rho_\theta\left((k+1)h - s\right) = \sum_{j=1}^{m} \alpha_j(s)\rho_\theta\left((k+1-j)h - s\right)\cdot$$

Remark 3.2.5. *Observe that if $\rho_\theta(\cdot)$ is a function which the respective discretization has a recursive form, then $\alpha_j(s) \equiv \alpha_j$.*

Example 3.2.3 satisfies the condition imposed in Assumption 3.2.4.

The next result establishes that the decomposition in Assumption 3.2.4 allows us to write the process as an *autoregressive process* of order m in which the error also depends on the functions $\alpha_j(s)$. This theorem, along with the previous example, generalizes Stein, Lopes and Medino [62, Example 4.1 and Propositions 4.1 and 4.2].

Theorem 3.2.6. *Let $\mathbf{X} = \{X(t); t \geq 0\}$ be defined by (3.2) with $\rho_\theta(\cdot)$ satisfying the Assumption 3.2.4. One discretization form for this process is given by*

$$X((k+1)h) = \sum_{j=1}^{m} \alpha_j(0) X((k+1-j)h) + \xi_{k,h},$$

where h is the discretization step size and

$$\begin{aligned} \xi_{k,h} &= \sum_{j=1}^{m} \left[\int_0^{(k+1)h} \alpha_j(s) \rho_\theta((k+1-j)h - s)\, dL(s) \right. \\ &- \left. \alpha_j(0) \int_0^{(k+1-j)h} \rho_\theta((k+1-j)h - s)\, dL(s) \right] \cdot \end{aligned} \tag{3.11}$$

Note that, up to now, we have used the explicit form of Kannan's solution (3.2) only in Proposition 2.2.4, Theorems 3.2.1 and 3.2.6. The estimations and the asymptotic study of the estimator (Chapter 2) depend only on the GLE

$$\begin{cases} dX(t) &= b(\theta, X_t)dt + dL(t), \quad t > 0 \\ X(t) &= X_0(t), \quad t \in [t_0, 0] \end{cases} \tag{3.12}$$

where $b(\theta, X_t)$ is defined as

$$b(\theta, X_t) = -\int_0^t X(s) \mu_{\theta,t}(ds) \cdot$$

We conclude this section highlighting that, as in Alcântara [1] and Stein, Lopes and Medino [62], adding an $\alpha-$stable hypothesis to the Lévy process $\mathbf{L}$ will help us determine the distribution of error ξ_k. A very practical result from the point of view of numerical simulations, since instead of working with the exact form of the error, we can just simulate its distribution.

Proposition 3.2.7. *In addition to the hypothesis of Theorem 3.2.6, if $\mathbf{L}$ is a symmetric $\alpha-$stable Lévy process with $1 < \alpha \leq 2$, we have*

$$\xi_{k,h} \sim S_\alpha \left(\left(\sum_{u=0}^{m} \int_{E_{k,u}} |g_{k,u}(s)|^\alpha ds \right)^{1/\alpha}, 0, 0 \right), \tag{3.13}$$

where

$$g_{k,u}(s) = \sum_{j=1}^{m} \tilde{\alpha}_{j,u}(s) \rho_\theta((k+1-j)h - s),$$

$$\tilde{\alpha}_{j,u}(s) = \begin{cases} \alpha_j(s) - \alpha_j(0), & j \in \{0, \cdots, m-u\} \text{ and } u \in \{1, \cdots, m-1\}, \\ \alpha_j(0), & j \in \{m-u+1, \cdots, m\} \text{ and } u \in \{1, \cdots, m\}, \end{cases}$$

and $E_{k,u} = \big[(k+1-u)h, (k-u))h\big)$, for $u = 0, \cdots, m-1$, and $E_{k,m} = \big[0, (k+1-m)h\big)$.

Remark 3.2.8. *The restriction $1 < \alpha \leq 2$ in the Proposition 3.2.7 is to guarantee that (3.2) is a solution of (3.12) in the sense of Santos [59, Theorem 2.3.1].*

3.3 A Modified FMLE (mFMLE) and Simulation Results

Consider discrete times $t_k = (k-1)h$, for $k = 1, 2, \cdots$ and a fixed positive size of discretization $h > 0$. Let $X(t_1), \cdots, X(t_n)$ be a discrete time observed path of a process $\mathbf{X}$ which is solution of a GLE (3.12). Denote all the past information of the process $\mathbf{X}$ until the time t_k by X_{t_k}, that is

$$X_{t_k} = \{X(t_1), \cdots, X(t_k)\}\cdot$$

Consider a discretization of the likelihood function

$$l(\theta, X_t) = \frac{1}{\sigma^2}\int_0^t b(\theta, X_s)dX^c(s) - \frac{1}{2\sigma^2}\int_0^t b^2(\theta, X_s)ds$$

using the jump filtering technique to approximate the increments of the continuous martingale part X^c, which was described in the Section 2.4, that is

$$l(\theta, X_{t_n}) = \left\{ \frac{1}{\sigma^2}\sum_{i=1}^n b(\theta, X_{t_i})\Delta_i X \mathbf{1}_{[|\Delta_i X| \leq v_n^i]} - \frac{1}{2\sigma^2}\sum_{i=1}^n b^2(\theta, X_{t_i})\Delta_i \right\}\cdot \tag{3.14}$$

Observe that a discretization of the GLE (3.12) satisfies

$$\Delta_k X = b(\theta_0, X_{t_k})h + \Delta_k L$$

which implies that $b(\theta_0, X_{t_k}) = \frac{1}{h}(\Delta_k L - \Delta_k X)$. Applying this expression to the filtered likelihood function (3.14), we approximate $l(\theta_0, X_{t_n})$. Then, we can define a *modified FMLE* (mFMLE) by

$$\hat{\theta}_T^{mFMLE} = \arg\min_{\theta\in\Theta_0} |l(\theta, X_{t_n}) - l(\theta_0, X_{t_n})|\,, \tag{3.15}$$

where $\{t_1, \cdots, t_n\}$ is a discretization of $[0, T]$.

Note that the FMLE $\hat{\theta}_T^{FMLE}$ defined in (3.1) can be written in function of (3.14) as

$$\hat{\theta}_T^{FMLE} = \arg\max_{\theta\in\Theta_0} l(\theta, X_{t_n})\cdot$$

Tables 3.1 and 3.2 present mean and standard deviation of the FMLE $\hat{\theta}_{100}^{FMLE}$ and mFMLE $\hat{\theta}_{100}^{mFMLE}$ for 100 Monte Carlo simulations of the GOU-FE with $n = 1000$ and $\sigma = 1$. The difference between them is that in the first table we use the same Lévy process generating X_T and getting the estimates $\hat{\theta}_T^{mFMLE}$, while in the second we use two independent paths but with the same distribution law.

To generate Table 3.1, in each line, we follow the steps:

1. fix the parameter $\theta^\top = (\theta_1, \theta_2, \theta_3)$ in the first column;
2. fix the search region Θ_0 for the heuristics FMLE and mFMLE;
3. simulated 100 Lévy process $\mathbf{L}$;
4. obtain the corresponding 3-parameter GOU-FE $\mathbf{X}$;
5. estimate the parameter θ in each simulation by FMLE and mFMLE;
6. compute the mean and standard deviation (std dv) of the estimations from both estimators.

Table 3.1: Mean and Standard Deviation of FMLE $\hat{\theta}_{100}^{FMLE}$ and FMLEm $\hat{\theta}_{100}^{mFMLE}$ for 100 Monte Carlo simulations of the 3-parameter GOU-FE with $n = 1000$ and $\sigma = 1$, for the same Lévy process generating X_T and obtaining $\hat{\theta}_T^{mFMLE}$.

$\theta^\top$	mean($\hat{\theta}_{100}^{FMLE}$)	mean($\hat{\theta}_{100}^{mFMLE}$)
(0.5, 0.2, 0.1)	(0.5330, 0.2930, 0.0895)	(0.5020, 0.1990, 0.1070)
(0.5, 0.2, 0.2)	(0.4920, 0.2700, 0.1170)	(0.5020, 0.1990, 0.2060)
(0,7, 0.2, 0.1)	(0.6840, 0.2935, 0.0835)	(0.7000, 0.1990, 0.1070)
(0.7, 0.2, 0.2)	(0.6350, 0.2780, 0.1140)	(0.7000, 0.1990, 0.2060)
(0.5, 0.3, 0.1)	(0.5365, 0.3155, 0.1110)	(0.5020, 0.2980, 0.1070)
(0.5, 0.3, 0.2)	(0.5335, 0.3245, 0.1690)	(0.5020, 0.2980, 0.2060)
(0,7, 0.3, 0.1)	(0.6920, 0.3100, 0.0955)	(0.7000, 0.2980, 0.1070)

$\theta^\top$	std dv($\hat{\theta}_{100}^{FMLE}$)	std dv($\hat{\theta}_{100}^{mFMLE}$)
(0.5, 0.2, 0.1)	(0.1266, 0.1971, 0.1095)	(0.0200, 0.0100, 0.0700)
(0.5, 0.2, 0.2)	(0.1584, 0.1002, 0.1168)	(0.0200, 0.0100, 0.0600)
(0,7, 0.2, 0.1)	(0.1257, 0.2277, 0.1018)	(0.0000, 0.0100, 0.0700)
(0.7, 0.2, 0.2)	(0.1582, 0.0954, 0.1204)	(0.0000, 0.0100, 0.0600)
(0.5, 0.3, 0.1)	(0.1489, 0.1226, 0.1169)	(0.0200, 0.0200, 0.0700)
(0.5, 0.3, 0.2)	(0.1498, 0.0463, 0.0600)	(0.1081, 0.0200, 0.0200)
(0,7, 0.3, 0.1)	(0.1306, 0.0959, 0.1018)	(0.0000, 0.0200, 0.0700)

Table 3.2: Mean and Standard Deviation of FMLE $\hat{\theta}_{100}^{FMLE}$ and mFMLE $\hat{\theta}_{100}^{mFMLE}$ for 100 Monte Carlo simulations of the 3-parameter GOU-FE with $n = 1000$ and $\sigma = 1$, for the equal in law Lévy process generating X_T and obtaining $\hat{\theta}_T^{mFMLE}$.

$\theta^\top$	mean($\hat{\theta}_{100}^{FMLE}$)	mean($\hat{\theta}_{100}^{mFMLE}$)
(0.5, 0.2, 0.1)	(0.4905, 0.2635, 0.0985)	(0.5135, 0.7785, 0.6965)
(0.5, 0.2, 0.2)	(0.5005, 0.2395, 0.1355)	(0.4930, 0.5850, 0.4340)
(0,7, 0.2, 0.1)	(0.7500, 0.1445, 0.1540)	(0.5750, 0.6765, 0.6325)
(0.7, 0.2, 0.2)	(0.6255, 0.2355, 0.1510)	(0.5400, 0.6865, 0.5175)
(0.5, 0.3, 0.1)	(0.4650, 0.3245, 0.1175)	(0.5565, 0.7310, 0.6015)
(0.5, 0.3, 0.2)	(0.4425, 0.3010, 0.2340)	(0.5665, 0.5490, 0.3845)

$\theta^\top$	std dv($\hat{\theta}_{100}^{FMLE}$)	std dv($\hat{\theta}_{100}^{mFMLE}$)
(0.5, 0.2, 0.1)	(0.1205, 0.1977, 0.1149)	(0.289, 0.1811, 0.1862)
(0.5, 0.2, 0.2)	(0.1413, 0.0509, 0.1138)	(0.2648, 0.2351, 0.2590)
(0,7, 0.2, 0.1)	(0.1393, 0.0550, 0.1084)	(0.2681, 0.1852, 0.2229)
(0.7, 0.2, 0.2)	(0.1250, 0.0416, 0.0992)	(0.2729, 0.2106, 0.2396)
(0.5, 0.3, 0.1)	(0.1658, 0.1069, 0.1406)	(0.2973, 0.1609, 0.2287)
(0.5, 0.3, 0.2)	(0.1928, 0.0882, 0.1660)	(0.2730, 0.3007, 0.2748)

The generation of Table 3.2 is a little different from Table 3.1. The following steps are followed:

1. fix the parameter $\theta^{\top} = (\theta_1, \theta_2, \theta_3)$ in the first column;
2. fix the search region Θ_0 for the heuristics FMLE and mFMLE;
3. simulated 100 Lévy process $\mathbf{L}$;
4. obtain the corresponding 3-parameter GOU-FE $\mathbf{X}$;
5. simulated other 100 Lévy process $\mathbf{L}$ with same distribution of that in 3;
6. estimate the parameter θ in each simulation by FMLE and mFMLE (using the simulated Lévy processes from 5);
7. compute the mean and standard deviation (std dv) of the estimations from both estimators.

Just like Table 2.1 in Chapter 2, Tables 3.1 and 3.2 were generated through the language and environment for statistical computing R and compiled on a server with 32 threads, in addition to using parallel computing. Each line required an average of 18 computing hours. Therefore, we reduced the number of simulations per line, from 500 in Table 2.1 to 100 in Tables 3.1 and 3.2.

For the fixed parameters, Table 3.1 shows that the second estimator $\hat{\theta}_{100}^{mFMLE}$ obtains a better performance with a more precise estimation ($|\theta_j - \hat{\theta}_{100,j}^{FMLE}| < |\theta_j - \hat{\theta}_{100,j}^{mFMLE}|$) and with less variability ($std\ dv\ \hat{\theta}_{100,j}^{FMLE} < std\ dv\ \hat{\theta}_{100,j}^{mFMLE}$). This was expected since removing the dependence of the SDE by the noise would facilitate obtaining the best parameter to model the time series.

For Table 3.2, the choice of parameters was kept and the difference was to take independent Lévy paths to simulate the process and to obtain the estimator $\hat{\theta}_T^{mFMLE}$. Despite taking the same law to generate the two Levy's processes, we observe that $\hat{\theta}_T^{FMLE}$ has better results than $\hat{\theta}_T^{mFMLE}$, for the choices of parameters taken. This indicates that $\hat{\theta}_T^{mFMLE}$ can perform better only if we have a very close approximation of the noise and also a good idea of the law of the noise. It can not be so useful when our time series is generated by a 3-parameter GOU-FE process. Actually, only the estimation of the first parameter θ_1 is good.

The results presented in Tables 2.1, 3.1 and 3.2 were obtained through local searches. We emphasize that for global searches, optimization methodologies can be used. We can use *Variable Neighborhood Search* (VNS) or even global optimization methodologies (as in Dorea [10]).

3.4 Proofs

Proof of the Theorem 3.2.1

Denote by $\mathcal{T}(f)$ the Laplace transform of a given function f.

Proof of the Theorem 3.2.1. Set $\mu_{\theta,t}(ds) = \gamma(t-s)ds$. As in Alcântara [1] and Santos [59], applying the Laplace transform in (3.3) we obtain the memory function of a GLE from

$$\mathcal{T}(\gamma) = -\frac{\mathcal{T}(\rho')}{\mathcal{T}(\rho)}. \tag{3.16}$$

Note that, for s which the following Laplace transforms are well defined,

$$
\begin{aligned}
\mathcal{T}(\rho) &= \mathcal{T}\left((1-\theta_3)e^{-\theta_1 t} + \theta_3\cos(\theta_2 t)\right) \\
&= (1-\theta_3)\frac{1}{s+\theta_1} + \theta_3\frac{s}{s^2+\theta_2^2} \\
&= \frac{(1-\theta_3)(s^2+\theta_2^2) + \theta_3(s+\theta_1)s}{(s+\theta_1)(s^2+\theta_2^2)} \\
&= \frac{\left(s+\frac{\theta_1\theta_3}{2}\right)^2 - \left(\frac{\theta_1\theta_3}{2}\right)^2 + \theta_2^2(1-\theta_3)}{(s+\theta_1)(s^2+\theta_2^2)} \\
&= \frac{(s+\nu_0)^2 + \nu_1^2 sign(\nu)}{(s+\theta_1)(s^2+\theta_2^2)} \qquad (3.17)
\end{aligned}
$$

where $\nu_0 = \theta_1\theta_3/2$, $\nu = -(\theta_1\theta_3/2)^2 + \theta_2^2(1-\theta_3)$ and $\nu_1^2 sign(\nu) = \nu$. Observe that $\nu > 0$, $\nu < 0$ or $\nu = 0$, gives us $\nu_1^2 = \nu$, $\nu_1^2 = -\nu$ or $\nu_1^2 = 0$, respectively.

Applying Laplace transform to $\rho_\theta(t)$, we obtain

$$
\begin{aligned}
\mathcal{T}(\rho') &= \mathcal{T}\left(-\theta_1(1-\theta_3)e^{-\theta_1 t} - \theta_2\theta_3\sin(\theta_2 t)\right) \\
&= -\frac{\theta_1(1-\theta_3)(s^2+\theta_2^2) + \theta_2^2\theta_3(s+\theta_1)}{(s+\theta_1)(s^2+\theta_2^2)} \\
&= -\frac{\theta_1(1-\theta_3)s^2 + \theta_2^2\theta_3 s + \theta_1(1-\theta_3)\theta_2^2 + \theta_1\theta_2^2\theta_3}{(s+\theta_1)(s^2+\theta_2^2)}. \qquad (3.18)
\end{aligned}
$$

Replacing (3.17) and (3.18) in (3.16), we obtain

$$
\begin{aligned}
\mathcal{T}(\gamma) &= \frac{\theta_1(1-\theta_3)s^2 + \theta_2^2\theta_3 s + \theta_1(1-\theta_3)\theta_2^2 + \theta_1\theta_2^2\theta_3}{(s+\nu_0)^2 + \nu_1^2 sign(\nu)} \\
&= \theta_1(1-\theta_3)\frac{s^2}{(s+\nu_0)^2+\nu_1^2 sign(\nu)} + \theta_2^2\theta_3\frac{s}{(s+\nu_0)^2+\nu_1^2 sign(\nu)} \\
&+ \theta_1\theta_2^2\frac{1}{(s+\nu_0)^2+\nu_1^2 sign(\nu)}.
\end{aligned}
$$

Using that

$$
s^2 = (s+\nu_0)^2 + \nu_1^2 sign(\nu) - 2\nu_0 s - \nu_0^2 - \nu_1^2 sign(\nu),
$$

we obtain

$$
\begin{aligned}
\mathcal{T}(\gamma) &= \theta_1(1-\theta_3)\left[1 - \frac{2\nu_0 s + \nu_0^2 + \nu_1^2 sign(\nu)}{(s+\nu_0)^2+\nu_1^2 sign(\nu)}\right] \\
&+ \theta_2^2\theta_3\frac{s}{(s+\nu_0)^2+\nu_1^2 sign(\nu)} \\
&+ \theta_1\theta_2^2\frac{1}{(s+\nu_0)^2+\nu_1^2 sign(\nu)} \\
&= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\frac{s+\nu_0}{(s+\nu_0)^2+\nu_1^2 sign(\nu)} \\
&+ \tilde{\kappa}(\theta)\frac{1}{(s+\nu_0)^2+\nu_1^2 sign(\nu)}, \qquad (3.19)
\end{aligned}
$$

where

$$
\begin{aligned}
\tilde{\kappa}(\theta) &= \theta_1\theta_2^2 - \nu_0\left(\theta_2^2\theta_3 - \theta_1(1-\theta_3)2\nu_0\right) - \theta_1(1-\theta_3)\left(\nu_0^2 + \nu_1^2 sign(\nu)\right), \\
\kappa_1(\theta) &= \theta_2^2\theta_3 - \theta_1(1-\theta_3)2\nu_0.
\end{aligned}
$$

Now, we need to analyse the three possible values of the $sign(\nu)$.

Case $\nu > 0$ **(**$sign(\nu) = 1$**).** Then, $\nu_1^2 = \nu$ and (3.19) imply

$$\begin{aligned}\mathcal{T}(\gamma) &= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\frac{s+\nu_0}{(s+\nu_0)^2+\nu^2} + \frac{\tilde{\kappa}(\theta)}{\nu}\frac{\nu}{(s+\nu_0)^2+\nu^2} \\ &= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\mathcal{T}\left(e^{-\nu_0 t}\cos(\nu t)\right) + \kappa_{2,1}(\theta)\mathcal{T}\left(e^{-\nu_0 t}\sin(\nu t)\right),\end{aligned}$$

where $\kappa_{2,1} = \frac{1}{\nu}\left(\theta_1\theta_2^2 - \nu_0\left(\theta_2^2\theta_3 - \theta_1(1-\theta_3)2\nu_0\right) - \theta_1(1-\theta_3)\left(\nu_0^2+\nu\right)\right)$.

Case $\nu < 0$ **(**$sign(\nu) = -1$**).** Then, $\nu_1^2 = -\nu$ and (3.19) imply

$$\begin{aligned}\mathcal{T}(\gamma) &= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\frac{s+\nu_0}{(s+\nu_0)^2-\nu^2} + \frac{\tilde{\kappa}(\theta)}{\nu}\frac{\nu}{(s+\nu_0)^2-\nu^2} \\ &= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\mathcal{T}\left(e^{-\nu_0 t}\cosh(\nu t)\right) + \kappa_{2,-1}(\theta)\mathcal{T}\left(e^{-\nu_0 t}\sinh(\nu t)\right),\end{aligned}$$

in which $\kappa_{2,-1} = \frac{1}{\nu}\left(\theta_1\theta_2^2 - \nu_0\left(\theta_2^2\theta_3 - \theta_1(1-\theta_3)2\nu_0\right) - \theta_1(1-\theta_3)\left(\nu_0^2-\nu\right)\right)$.

Case $\nu = 0$ **(**$sign(\nu) = 0$**).** Then, $\nu_1^2 = 0$ and (3.19) imply

$$\begin{aligned}\mathcal{T}(\gamma) &= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\frac{s+\nu_0}{(s+\nu_0)^2} + \tilde{\kappa}(\theta)\frac{1}{(s+\nu_0)^2} \\ &= \theta_1(1-\theta_3)\mathcal{T}(\delta_0) + \kappa_1(\theta)\mathcal{T}\left(e^{-\nu_0 t}\right) + \kappa_{2,-1}(\theta)\mathcal{T}\left(te^{-\nu_0 t}\right),\end{aligned}$$

where $\kappa_{2,0} = \theta_1\theta_2^2 - \nu_0\left(\theta_2^2\theta_3 - \theta_1(1-\theta_3)2\nu_0\right) - \theta_1(1-\theta_3)\nu_0^2$.

The theorem follows from linearity of $\mathcal{T}$ and the inverse Laplace transform $\mathcal{T}^{-1}$.

∎

Proof of the Proposition 3.2.2

Proof of the Proposition 3.2.2. It follows from (3.6) and (3.7) that

$$\begin{aligned}X((k+1)h) &= (1-\theta_3)A((k+1)h) + \theta_3 B((k+1)h) \\ &= (1-\theta_3)\left(e^{-\theta_1 h}A(kh) + \varepsilon_{A,k+1}\right) \\ &+ \theta_3\left(2\cos(\theta_2 h)B(kh) - B((k-1)h) + \varepsilon_{B,k+1}\right).\end{aligned}$$

By adding and subtracting appropriate multiples of $A(kh)$ and $B(kh)$, we obtain

$$\begin{aligned}X((k+1)h) &= (1-\theta_3)\left(e^{-\theta_1 h} + 2\cos(\theta_2 h)\right)A(kh) - (1-\theta_3)2\cos(\theta_2 h)A(kh) \\ &+ \theta_3\left(e^{-\theta_1 h} + 2\cos(\theta_2 h)\right)B(kh) - \theta_3 e^{-\theta_1 h}B(kh) \\ &- \theta_3 B((k-1)h) + (1-\theta_3)\varepsilon_{A,k+1} + \theta_3\varepsilon_{B,k+1} \\ &= \left(e^{-\theta_1 h} + 2\cos(\theta_2 h)\right)X(kh) - (1-\theta_3)2\cos(\theta_2 h)A(kh) \\ &- \theta_3 e^{-\theta_1 h}B(kh) - \theta_3 B((k-1)h) + (1-\theta_3)\varepsilon_{A,k+1} + \theta_3\varepsilon_{B,k+1}.\end{aligned}$$

Denoting $\phi_1 = e^{-\theta_1 h} + 2\cos(\theta_2 h)$. Applying (3.6) and (3.7) for $A(kh)$ and $B(kh)$, we have

$$\begin{aligned}X((k+1)h) &= \phi_1 X(kh) - (1-\theta_3)2\cos(\theta_2 h)\left(e^{-\theta_1 h}A((k-1)h) + \varepsilon_{A,k}\right) \\ &- \theta_3 e^{-\theta_1 h}\left(2\cos(\theta_2 h)B((k-1)h) - B((k-2)h) + \varepsilon_{B,k}\right) \\ &- \theta_3 B((k-1)h) + (1-\theta_3)\varepsilon_{A,k+1} + \theta_3\varepsilon_{B,k+1} \\ &= \phi_1 X(kh) - 2\cos(\theta_2 h)e^{-\theta_1 h}\left((1-\theta_3)A((k-1)h) + \theta_3 B((k-1)h)\right) \\ &- (1-\theta_3)2\cos(\theta_2 h)\varepsilon_{A,k} - \theta_3 e^{-\theta_1 h}\varepsilon_{B,k} + (1-\theta_3)\varepsilon_{A,k+1} + \theta_3\varepsilon_{B,k+1} \\ &- \theta_3 B((k-1)h) + \theta_3 e^{-\theta_1 h}B((k-2)h).\end{aligned} \tag{3.20}$$

Observe, from (3.6), that

$$(1-\theta_3)A((k-1)h) - (1-\theta_3)e^{-\theta_1 h}A((k-2)h) = (1-\theta_3)\varepsilon_{A,k-1},$$

which implies

$$\begin{aligned}
-\theta_3 B((k-1)h) + \theta_3 e^{-\theta_1 h}B((k-2)h) &= -X((k-1)h) + (1-\theta_3)A((k-1)h) \\
&+ e^{-\theta_1 h}X((k-2)h) - (1-\theta_3)e^{-\theta_1 h}A((k-2)h) \\
&= -X((k-1)h) + e^{-\theta_1 h}X((k-2)h) \\
&+ (1-\theta_3)\varepsilon_{A,k-1}, \qquad (3.21)
\end{aligned}$$

Applying (3.21) in (3.20), we show that

$$\begin{aligned}
X((k+1)h) &= \phi_1 X(kh) + \phi_2 X((k-1)h) + \phi_3 X((k-2)h) \\
&+ (1-\theta_3)\varepsilon_{A,k+1} - (1-\theta_3)2\cos(\theta_2 h)\varepsilon_{A,k} + (1-\theta_3)\varepsilon_{A,k-1} \\
&+ \theta_3\varepsilon_{B,k+1} - \theta_3 e^{-\theta_1 h}\varepsilon_{B,k}
\end{aligned}$$

where $\phi_2 = -1 - 2\cos(\theta_2 h)e^{-\theta_1 h}$ and $\phi_3 = e^{-\theta_1 h}$, which concludes the proof.

■

Proof of the Theorem 3.2.6

Proof of the Theorem 3.2.6. By the Equation (3.2) and the Assumption 3.2.4 we have

$$X_0 \sum_{j=1}^{m} \alpha_j(0)\rho_\theta\left((k+1-j)h\right) = X\left((k+1)h\right) - \sum_{j=1}^{m}\int_0^{(k+1)h} \alpha_j(s)\rho_\theta\left((k+1-j)h - s\right)dL(s). \qquad (3.22)$$

For $k \in \mathbb{N}$ and $j \in \{1, \cdots, k+1\}$, by (3.2) we obtain

$$X_0\rho_\theta\left((k+1-j)h\right) = X\left((k+1-j)h\right) - \int_0^{(k+1-j)h} \rho_\theta((k+1-j)h - s)dL(s),$$

which implies

$$\begin{aligned}
X_0 \sum_{j=1}^{m} \alpha_j(0)\rho_\theta\left((k+1-j)h\right) &= \sum_{j=1}^{m} \alpha_j(0)\left[X\left((k+1-j)h\right)\right. \\
&- \left.\int_0^{(k+1-j)h} \rho_\theta((k+1-j)h - s)dL(s)\right]. \qquad (3.23)
\end{aligned}$$

It follows from (3.22) and (3.23) that

$$\begin{aligned}
X\left((k+1)h\right) &= \sum_{j=1}^{m} \alpha_j(0)\left[X\left((k+1-j)h\right) - \int_0^{(k+1-j)h} \rho_\theta((k+1-j)h - s)dL(s)\right] \\
&+ \sum_{j=1}^{m}\int_0^{(k+1)h} \alpha_j(s)\rho_\theta\left((k+1)h - s\right)dL(s),
\end{aligned}$$

which proves the theorem.

■

Proof of the Proposition 3.2.7

Denote the stable distributions by $S_\alpha(\sigma,\beta,\mu)$ which the characteristic function (cf. Samorodnitsky and Taqqw [58]) is determined by

$$\alpha\in(0,2],\ \sigma\geq 0,\ \beta\in[-1,1],\mu\in\mathbb{R}.$$

We recall that a Lévy process $\mathbf{L}=\{L(t);t\geq 0\}$ is $\alpha-$*stable*, $0<\alpha\leq 2$ if

$$L(t)-L(s)\sim S_\alpha\left((t-s)^{1/\alpha},\beta,0\right).$$

In the case where $\beta=0$, we say that $\mathbf{L}$ is *symmetric*.

From Samorodnitsky and Taqqw [58, Property 3.2.2], we know that if $f\in L^\alpha(E)$, then $\int_E f(s)dL(s)\sim S_\alpha(\sigma_f,\beta_f,\mu_f)$, where

$$\begin{aligned}\sigma_f &= \left(\int_E |f(x)|^\alpha dx\right)^{1/\alpha},\\ \beta_f &= \frac{\beta\int_E |f(x)|^\alpha sign(f(x))dx}{\int_E |f(x)|^\alpha dx},\\ \mu_f &= \begin{cases} 0, & \alpha\neq 1,\\ -\frac{2}{\pi}\beta\int_E f(x)\ln|f(x)|dx, & \alpha=1\cdot\end{cases}\end{aligned}$$

By a natural extension of the Samorodnitsky and Taqqw [58, Property 1.2.1], we can prove, using characteristic functions, that if $X_1,\cdots,X_{m+1}$ are independent random variables with $X_u\sim S_\alpha(\sigma_u,\beta_u,\mu_u)$, $u=1,\cdots,m+1$, then

$$\sum_{u=1}^{m+1}X_u\sim S_\alpha\left(\left(\sum_{u=1}^{m+1}\sigma_u^\alpha\right)^{1/\alpha},\frac{\sum_{u=1}^{m+1}\sigma_u^\alpha\beta_u}{\sum_{u=1}^{m+1}\sigma_u^\alpha},\sum_{u=1}^{m+1}\mu_u\right)\cdot \tag{3.24}$$

Proof of Proposition 3.2.7. By (3.11) we are able to write

$$\begin{aligned}\xi_{k,h} &= \sum_{j=1}^{m}\left[\int_0^{(k+1-j)h}(\alpha_j(s)-\alpha_j(0))\,\rho((k+1-j)h-s)dL(s)\right.\\ &+ \left.\int_{(k+1-j)h}^{(k+1)h}\alpha_j(s)\rho((k+1-j)h-s)dL(s)\right]\\ &= \int_0^{(k+1-m)h}\sum_{j=1}^{m}\alpha_j(0)\rho((k+1-j)h-s)dL(s)\\ &+ \sum_{u=0}^{m-1}\int_{(k-u)h}^{(k+1-u)h}\left(\sum_{j=1}^{m-u}(\alpha_j(s)-\alpha_j(0))\,\rho((k+1-j)h-s)\right.\\ &+ \left.\sum_{j=m+1-u}^{m}\alpha_j(s)\rho((k+1-j)h-s)\right)dL(s)\\ &= \sum_{u=0}^{m}\int_{E_{k,u}}g_{k,u}(s)dL(s)\cdot\end{aligned}$$

Observe that $E_{k,u}$ are disjoint sets. The independent increments of the Lévy process imply that $\xi_{k,h}$ is a sum of $m+1$ independent random variables. Then, by Samorodnitsky and Taqqw [58, Property 3.2.2],

$$\int_{E_{k,u}} g_{k,u}(s)dL(s) \sim S_\alpha\left(\left(\int_{E_{k,u}} |g_{k,u}(s)|^\alpha ds\right)^{1/\alpha}, 0, 0\right).$$

Therefore, the extension of the Samorodnitsky and Taqqw [58, Property 1.2.1] given in (3.24) implies (3.13).

■

Chapter 4

Asymptotics for MLE and FMLE: a Particular Class

4.1 Introduction

In previous chapters we have studied the problem of MLE for the GLE,

$$\begin{cases} dX(t) &= b(\theta, X_t)dt + dL(t), \;\; t > 0 \\ X(t) &= X_0(t), \;\; t \in [0, T]. \end{cases} \tag{4.1}$$

Asymptotics for the MLE $\hat{\theta}(t)$ were obtained and discrete approximations $\hat{\theta}_T^{FMLE}$ and $\hat{\theta}_T^{mFMLE}$ were proposed. Though a theoretical study of the discretized estimators were not carried out, its behavior was evaluated by simulations.

In this chapter, we consider a particular class of GLE for which the drift satisfies

$$b(\theta, X_t) = -\int_0^t X(s)\mu_{\theta,t}(ds) \;\text{ and }\; \mu_{\theta,t} = \sum_{j=1}^N \theta_j \mu_t^{(j)}, \tag{4.2}$$

where the signed measures $\mu_t^{(j)}$ are finite at intervals $[0, t]$. This restriction appears naturally when we analyze the OU, cosine and stochastic delay (SDDE) processes. For this type of processes the parameter θ can be linearly separated and depends exclusively on the past history of the process $\mathbf{X}$ until say time t_k, that is, $X_{t_k} = \{X(t_1), \cdots, X(t_k)\}$. Moreover, for this class of GLE the Radon-Nikodym density of P_θ with respect to P_0 takes up a much simpler form and we can derive an explicit expression for the MLE,

$$\hat{\theta}(t) := \arg\max_{\theta \in \Theta_0} \frac{dP_\theta^t}{dP_0^t}(X_t) = S(t)^{-1}A(t),$$

where the matrices $S(t)$ and $A(t)$ are properly defined functions of the drift function $b(\theta, X_t)$ and the continuous martingale part X^c. Based on this and under much milder assumptions than those of Theorem 2.3.4, our Theorem 4.2.4 establishes for $\hat{\theta}(t)$ the strong consistency, asymptotic normality and asymptotic efficiency in the sense of Hájek-Le Cam Convolution Theorem. Its proof is less technical and basically requires a version of the CLT for continuous time martingales.

In Section 4.3, we study the asymptotic behavior of the corresponding discretized and filtered approximation $\hat{\theta}_n^{FMLE} := S_n^{-1}A_n$ where S_n and A_n are approximations of $S(t)$ and $A(t)$ respectively.

As shown in Theorem 4.3.2, with a good convergence rate for some stochastic integrals approximations by Riemann sums, the FMLE will have the same asymptotic behavior as that of MLE.

In the last section we gather all the proofs. The main used tools are properties of Itô integrals, CLT and LLN for continuous time martingales (see, respectively, Klebaner [27], Küchler Sϕrensen [34], Crimaldi and Pratelli [8] and Liptser [38]).

4.2 An Explicit MLE for a Particular Class of GLE

In this section, we present a class of GLEs for which an explicit form for the MLE can be obtained. We start this section with few examples that induced us to consider the class of processes to be studied.

1. Langevin Equation:
$$dX(t) = -\theta X(t)dt + dL(t), \quad \theta \in \mathbb{R}_+; \tag{4.3}$$
2. Cosine Process:
$$dX(t) = -\theta^2 \int_0^t X(s)dsdt + dL(t), \theta \in \mathbb{R};$$
3. Stochastic Delay Differential Equation:
$$dX(t) = (\theta_1 X(t) + \theta_2 X(t-1))\, dt + dL(t), \quad \theta^\top = (\theta_1, \theta_2) \in \Theta \subset \mathbb{R}^2. \tag{4.4}$$

For these examples the parameter θ can be linearly separated and depends exclusively on the past history of the process $\mathbf{X}$ until say time t, that is, $X_t = \{X(s); 0 \leq s \leq t\}$ and the associated signed measures satisfy the following decomposition
$$\mu_{\theta,t} = \sum_{j=1}^{N} \theta_j \mu_t^{(j)}, \tag{4.5}$$
where $\theta^\top = (\theta_1, \cdots, \theta_N) \in \Theta \subset \mathbb{R}^N$ and $\mu_t^{(j)}$ are signed measures that are finite at intervals $[0, s]$, $0 \leq s \leq t$. We will restrict the studies to a family of models that satisfies the following assumption which will enable us obtain a explicit MLE for the drift parameter of the GLE.

Assumption 4.2.1. *Let us assume that the the process $b(\theta, X_t)$ defined in (4.2) has the following decomposition*
$$b(\theta, X_t) = \sum_{j=1}^{N} \theta_j b_j(X_t), \tag{4.6}$$
where $b_j(X_t) \in L^2(dP_\theta \times dt)$ and the functions $b_j(X_t)$ and $b_k(X_t)$ are different, for each pair $j, k = 1, \cdots, N$, $j \neq k$.

It follows from Proposition 2.2.2 and Assumption 4.2.1 that the likelihood function is a curved exponential family in the sense of Küchler and Sϕrensen [33] (see Definition A.6.2) and it is obtained in the following result.

Proposition 4.2.2. *Under conditions of the Proposition 2.2.2 and Assumption 4.2.1, the likelihood process (Radon-Nikodym density process of P_θ with respect to P_0) is given by*
$$\frac{dP_\theta^t}{dP_0^t}(X_t) = \exp\left\{\theta^\top A(t) - \frac{1}{2}\theta^\top S(t)\theta\right\}, \tag{4.7}$$

where

$$A(t) = (A_j(t))_{N\times 1} = \left(\frac{1}{\sigma^2}\int_0^t b_j(X_s)dX^c(s)\right)_{N\times 1}$$

and

$$S(t) = (S_{jk}(t))_{N\times N} = \left(\frac{1}{\sigma^2}\int_0^t b_j(X_s)b_k(X_s)ds\right)_{N\times N}. \tag{4.8}$$

Here X^c denotes the continuous martingale part of $\mathbf{X}$ under P_θ.

Proposition 4.2.2 states that, if $S(t)$ is an invertible matrix, the *maximum likelihood estimator* (MLE) $\hat{\theta}(t)$ of $\theta_0 \in \Theta_0$ is given explicitly by

$$\hat{\theta}(t) := \arg\max_{\theta\in\Theta_0} \frac{dP_\theta^t}{dP_0^t}(X_t) = S(t)^{-1}A(t)\cdot \tag{4.9}$$

The next condition ensures, in a sense, the ergodicity of the process. This condition assists us in proving the main result for $\hat{\theta}(t)$ defined by (4.9).

Assumption 4.2.3. *Assume that, for each $\theta \in \Theta_0$, there exists a deterministic positive definite $N\times N$-matrix $\Sigma(\theta) = (\Sigma_{jk}(\theta))_{N\times N}$ and a positive function $\overline{\varphi}(t)$ such that $\overline{\varphi}(t) \uparrow \infty$ as $t\to\infty$ and $S(t)$ defined on (4.8) satisfies*

$$\overline{\varphi}(t)^{-1}S(t) \to \Sigma(\theta) \;\; P_\theta - a.s. \text{ as } t \to \infty\cdot$$

Sufficient conditions for ergodicity of diffusion processes have been well studied (see, for example, Masuda [46, 47]). In general, integrability conditions on the Lévy measure ν and a restriction on the parameter space $\Theta_0 \subset \Theta$ are required. Tran [64] listed recent researches on studies of ergodicity for diffusion processes. An example of the Assumption 4.2.3 being satisfied is the OU process in which the convergence of $t^{-1}\int_0^t X^2(s)ds$ is guaranteed by ergodicity.

Magdziarz and Weron [41, Theorem 1] presented an ergodic theorem that can be applied for the GOU process $\mathbf{X}$, that is, taking a appropriated Lévy process for which the correlation function of $\mathbf{X}$ decays to zero as $t\to\infty$ and for a suitable f, the temporal and ensemble averages coincide, i.e.,

$$\frac{1}{T}\int_0^T f(X(t))dt \to \mathbf{E}[f],$$

provided that $\mathbf{E}[f] < \infty\cdot$ It becomes sufficient conditions for Assumption 4.2.3.

Now, we are able to establish that the statistical experiment $\{P_\theta; \theta\in\Theta\}$ satisfies the LAN property and the asymptotic behavior of the estimator: its strong consistency and the asymptotic normal distribution. The main tool to prove this result is the version of the CLT that we will prove in the Lemma 4.4.1.

The next result is a version of the Theorem 2.3.4, with less restrictive assumptions, but with the drift satisfying (4.6).

Theorem 4.2.4. *Under conditions of the Proposition 4.2.2, suppose that Assumption 4.2.3 holds.*

1. *The MLE (4.9) is strongly consistent, that means,*

$$\hat{\theta}(t) \to \theta_0, \;\; P_{\theta_0} - a.s. \;\; as \;\; t\to\infty.$$

Moreover,

$$Law\left(\overline{\varphi}(t)^{1/2}\left(\hat{\theta}(t) - \theta_0\right)\Big| P_{\theta_0}\right) \to \Sigma(\theta_0)^{-1/2}N(0, I_N), \;\; as \;\; t\to\infty, \tag{4.10}$$

where I_N denotes de $N\times N-$identity matrix.

2. *The statistical experiment $\{P_\theta; \theta \in \Theta_0\}$ is LAN at each $\theta \in int\ \Theta_0$ with the Fisher information matrix $\Sigma(\theta)$ and rate of convergence $\overline{\varphi}(t)^{-1/2}$. That means, for all $h \in \mathbb{R}^N$,*

$$Law\left(\log \frac{dP^t_{\theta+h\overline{\varphi}(t)^{-1/2}}}{dP^t_\theta}\middle| P_\theta\right) \to h^\top N\left(0, \Sigma(\theta)\right) - \frac{1}{2}h^\top\Sigma(\theta)h, \;\; as\;\; t \to \infty\cdot$$

Furthermore, the MLE is asymptotically efficient in the sense of Hájek-Le Cam Convolution Theorem.

The challenge now is to find a discretized form $\hat{\theta}_n$ of the estimator $\hat{\theta}(t)$ defined in (4.9) preserving a similar asymptotic behavior. This is important from the point of view of applications since, although the generating process $\mathbf{X}$ has paths in continuous time, in general, time series are observed in discrete times $t_1, \cdots, t_n$, even though the size of the intervals may be extremely small. This issue is covered in the next section.

4.3 Asymptotic Behavior of the FMLE

In this Section, we present the discrete time estimator for θ_0 via a sample $X_{t_1}, \cdots, X_{t_n}$ from a solution process of the GLE (4.1) with the restriction (4.6). We start by setting some notations and conditions and then present the main result of the estimation.

Let $\mathbf{L} = \{L(t); t \geq 0\}$ be a Lévy process with characteristic triplet $(0, \sigma^2, \nu)$. Suppose that the process $\mathbf{X} = \{X(t); t \geq 0\}$ solution of the GLE (4.1) was observed at discrete time points $0 = t_1 < \cdots < t_n$. We denote the true value of the parameter corresponding to the observed process path by $\theta_0^\top = (\theta_1^0, \cdots, \theta_N^0) \in \Theta_0$.

Consider that the jump component $\mathbf{J} = \{J(t); t \geq 0\}$ of the noise $\mathbf{L}$ is a compound Poisson process given by

$$J(t) = \sum_{j=1}^{N(t)} Y_j, \tag{4.11}$$

where the number of jumps is a Poisson process $\mathbf{N} = \{N(t); t \geq 0\}$ with intensity $\lambda = \nu(\mathbb{R}) < \infty$ and the jumps' size are independent and identically distributed random variables $Y_1, \cdots, Y_{N(t)}$ with common distribution F.

Set $\Delta_i := t_{i+1} - t_i$, for $i = 1, \cdots, n-1$, and $\Delta_n := \max\limits_{i=1,\cdots,n-1} \Delta_i$. Assume that $\Delta_n \to 0$ and $t_n \to \infty$ as $n \to \infty$. Given a real function g, we will denote the i-th increment of g as $\Delta_i g = g(t_{i+1}) - g(t_i)$. Denote by $X(t_i)$ and $X_{t_i} = \{X_{t_1}, \cdots, X_{t_i}\}$, respectively, the process observed at time t_i and the process' history until time t_i.

As mentioned in Sections 2.4 and 3.3 and inspired by the studies of threshold estimation for stochastic models with jumps done in Gloter, Loukianova and Mai [13], Mai [42, 43] and Mancini [44, 45] we propose the following *filtered maximum likelihood estimator* (FMLE)

$$\hat{\theta}_n^{FMLE} := S_n^{-1} A_n,$$

where

$$A_n = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1} b_j(X_{t_i})\Delta_i X \mathbf{1}_{\left[|\Delta_i X| \leq \Delta_n^\beta\right]}\right)_{N\times 1}$$

and

$$S_n = \left(\frac{1}{\sigma^2} \sum_{i=1}^{n-1} b_j(X_{t_i}) b_k(X_{t_i}) \Delta_i \right)_{N\times N}, \tag{4.12}$$

for a fixed $\beta \in (0, 1/2)$.

Observe that A_n and S_n are discrete time versions of $A(t)$ and $S(t)$ defined in the Proposition 4.2.2, respectively.

Denote by $\mathbf{D} = \{D(t); t \geq 0\}$ the drift component of $\mathbf{X}$ which, under Assumption 4.2.1 it can be written as

$$D(t) = \int_0^t b(\theta, X_s) ds = \sum_{j=1}^{N} \theta_j \int_0^t b_j(X_s) ds. \tag{4.13}$$

In order to determine the asymptotic behavior of the normalized estimator $\hat{\theta}_n^{FMLE}$, we assume the following conditions, in addition to those previously assumed in Section 4.2.

Assumption 4.3.1. *Assume that:*

1. $n\Delta_n t_n^{-1} = O(1)$, $t_n \Delta_n^{1-2\beta} = o(1)$ *and there exists* $\beta \in (0, 1/2)$ *such that the distribution* F *of the jump heights satisfies*
$$P_F\left(-2\Delta_n^\beta, 2\Delta_n^\beta\right) = o(t_n^{-1}), \quad as \;\; n \to \infty$$
where P_F *is the Lebesgue-Stieltjes measure associated with* F*;*

2. $\bar{\varphi}(t) = t$ *in Assumption 4.2.3;*

3. *under* P_{θ_0}, X^c, $\Delta_i W$, $\Delta_i D$ *and* $b_j(X_{t_i})$ *are mutually independent, for each* $i \in \{1, \cdots, n-1\}$ *and* $j \in \{1, \cdots, N\}$*;*

4. $$\max_{(j,k)\in\{1,\cdots,N\}^2} \left| \sum_{i=1}^{n-1} b_j(X_{t_i}) \int_{t_i}^{t_{i+1}} b_k(X_s) ds - \sum_{i=1}^{n-1} b_j(X_{t_i}) b_k(X_{t_i}) \Delta_i \right| = o_{P_{\theta_0}}\left(\sqrt{t_n}\right);$$

5. *the process* $\mathbf{X}$ *has finite second moments.*

The following result shows that $\hat{\theta}_n^{FMLE}$ has the same asymptotic distribution as MLE $\hat{\theta}(t)$ observed at continuous time.

Theorem 4.3.2. *In addition to the hypothesis of Theorem 4.2.4, assume that Assumption 4.3.1 holds. Then*

$$Law\left(\sqrt{t_n}\left(\hat{\theta}_n^{FMLE} - \theta_0\right)\Big| P_{\theta_0}\right) \to N(0, \Sigma(\theta_0)^{-1}), \;\; as \;\; n \to \infty.$$

Moreover, $\hat{\theta}_n^{FMLE}$ *is asymptotically efficient in the sense of Hájek-Le Cam Convolution Theorem.*

We finish this section by pointing out that the results obtained in this chapter generalize the results of Mai [42, 43] for the Langevin equation (4.3) and the Stochastic delay differential equation (4.4), but they cannot be used for the estimating the drift parameter of the GOU-FE (3.5).

4.4 Proofs

Proof of the Proposition 4.2.2

Proof of the Proposition 4.2.2. Consider $\theta' = 0 \in \Theta$. Under Assumption 4.2.1, $b(0, X_s) \equiv 0$ and the integral processes in (2.4) have the following decomposition

$$\int_0^t b(\theta, X_s)dX^c(s) = \sum_{j=1}^N \theta_j \int_0^t b_j(X_s)dX^c(s) \tag{4.14}$$

and

$$\int_0^t b^2(\theta, X_s)ds = \sum_{j=1}^N \sum_{k=1}^N \theta_j \theta_k \int_0^t b_j(X_s)b_k(X_s)ds. \tag{4.15}$$

Let $P_\theta^0 = P_0^0$ be the initial measure of the process $\mathbf{X}$. By (4.14), (4.15) and Proposition 2.2.2, the likelihood function can be written as

$$\frac{dP_\theta^t}{dP_0^t}(X_t) \quad = \quad \exp\left\{ \frac{1}{\sigma^2} \sum_{j=1}^N \theta_j \int_0^t b_j(X_s)dX^c(s) - \frac{1}{2\sigma^2} \sum_{j=1}^N \sum_{k=1}^N \theta_j \theta_k \int_0^t b_j(X_s)b_k(X_s)ds \right\},$$

which proves the proposition.

■

Proof of the Theorem 4.2.4

Under the Assumption 4.2.1, define a N-dimensional martingale $M(t)^\top = (M_1(t), \cdots, M_N(t))$ such that, for each $j \in \{1, \cdots, N\}$,

$$M_j(t) = \frac{1}{\sigma} \int_0^t b_j(X_s)dW(s), \tag{4.16}$$

where $\mathbf{W} = \{W(t); t \geq 0\}$ is a P_θ-Wiener process.

Thus, we are able to prove the following result, which will be an essential tool to prove the Theorem 4.2.4. Essentially, this result is a consequence of the CLT for $N-$dimensional martingales (cf. Theorem A.7.4, Crimaldi and Pratelli [8, Theorem 2.2], Küchler and Sϕrensen [33, Theorem A.7.7] or, [34, Theorem 2.1]).

Lemma 4.4.1. *Under Assumptions 4.2.1 and 4.2.3, we have that, for all $\theta \in \Theta_0$,*

$$Law\left(\overline{\varphi}(t)^{-1/2}M(t),\ \overline{\varphi}(t)^{-1}[M](t)\Big|\, P_\theta\right) \to \left(\Sigma(\theta)^{1/2}Z,\ \Sigma(\theta)\right), \quad as \ \ t \to \infty,$$

where Z is a $N-$dimensional random vector with standard normal distribution, $[M]$ denotes the quadratic covariation matrix of M and $\overline{\varphi}(t)$ and $\Sigma(\theta)$ are defined in Assumption 4.2.3.

Proof. By Assumption 4.2.3, applying Proposition A.7.3, we obtain that $M(t)$ is a continuous zero mean square integrable martingale with quadratic covariation matrix

$$[M](t) = \left(M_{jk}(t)\right)_{N\times N},$$

where

$$M_{jk}(t) = \frac{1}{\sigma^2} \int_0^t b_j(X_s)b_k(X_s)ds.$$

Clearly,

$$\sum_{j=1}^{N} |K_{jk}(t)|\, E_\theta \left[\sup_{s\le t} |\Delta M_k(s)|\right] \equiv 0,$$

where $K_{jk}(t) = \overline{\varphi}(t)^{-1/2}$ if $j = k$ or 0 otherwise. Moreover, under the Assumption 4.2.3,

$$\overline{\varphi}(t)^{-1} S(t) \to \Sigma(\theta),$$

P_θ-a.s. as $t \to \infty$. The result follows from the CLT for N-dimensional martingales A.7.4.

■

Remark 4.4.2. *The quadratic covariation matrix of $M(t)$ is the same matrix $S(t)$ that appears in (4.7).*

Proof of the Theorem 4.2.4. This proof will be divided into two steps.

Step 1. (Strong consistency and asymptotic normality of MLE $\hat{\theta}(t)$). It follows from the large of numbers law for martingales (see, for example, Theorem A.7.5) that the N-dimensional martingale $M(t)$ defined in (4.16) satisfies

$$\frac{1}{\overline{\varphi}(t)} M_j(t) = \frac{1}{\overline{\varphi}(t)} \int_0^t b_j(X_s) dW(s) \to 0, \;\; P_{\theta_0} - a.s. \;\text{ as }\; t \to \infty, \tag{4.17}$$

for each $j = 1, \cdots, N$. Observe that the continuous P_θ-martingale part $X^c(t)$ of $X(t)$ gives us

$$dX^c(s) = \sigma dW(s) + b\left(\theta, X_s\right) ds. \tag{4.18}$$

Under Assumption 4.2.1, we have for each $j = 1, \cdots, N$

$$M_j(t) \;=\; \frac{1}{\sigma^2} \int_0^t b_j\left(X_s\right) dX^c(s) - \frac{1}{\sigma^2} \sum_{k=1}^{N} \theta_k \int_0^t b_j\left(X_s\right) b_k\left(X_s\right) ds. \tag{4.19}$$

Using the decomposition (4.19), under θ_0, we can rewrite $A(t)$ as

$$A(t) = S(t)^\top \theta_0 + M(t), \tag{4.20}$$

where $M(t)$ is the N-dimensional martingale defined in (4.16). Hence, by (4.20),

$$\begin{aligned} \hat{\theta}(t) - \theta_0 \;&=\; S(t)^{-1} A(t) - \theta_0 \\ &=\; S(t)^{-1} \left(S(t)^\top \theta_0 + M(t)\right) - \theta_0 \\ &=\; S(t)^{-1} M(t) \\ &=\; \left(\frac{1}{\overline{\varphi}(t)} S(t)\right)^{-1} \frac{1}{\overline{\varphi}(t)} M(t), \end{aligned} \tag{4.21}$$

which converges to zero P_{θ_0}-a.s. as $t \to \infty$ due to Assumption 4.2.3 and (4.17).

Finally, by (4.21),

$$\overline{\varphi}(t)^{1/2} \left(\hat{\theta}(t) - \theta_0\right) = \left(\overline{\varphi}(t)[M](t)^{-1}\right) \overline{\varphi}(t)^{-1/2} M(t),$$

thus, Lemma 4.4.1 implies that (4.10) holds.

Step 2. (locally asymptotically normal property). For any $h^\top = (h_1, \cdots, h_N) \in \mathbb{R}^N$, observe that the Radon-Nikodym Chain Rule gives us

$$\log \frac{dP^t_{\theta + h\overline{\varphi}(t)^{-1/2}}}{dP^t_\theta} = l\left(\theta + h\overline{\varphi}(t)^{-1/2}, X_t\right) - l(\theta, X_t).$$

By Proposition 4.2.2,

$$\begin{aligned} l(\theta + h\overline{\varphi}(t)^{-1/2}, X_t) - l(\theta, X_t) &= \left(\theta + h\overline{\varphi}(t)^{-1/2}\right)^\top A(t) \\ &- \frac{1}{2}\left(\theta + h\overline{\varphi}(t)^{-1/2}\right)^\top S(t)\left(\theta + h\overline{\varphi}(t)^{-1/2}\right) \\ &- \theta^\top A(t) + \frac{1}{2}\theta^\top S(t)\theta \\ &= \left(\theta + h\overline{\varphi}(t)^{-1/2} - \theta\right)^\top A(t) \\ &- \frac{1}{2}\left(\theta + h\overline{\varphi}(t)^{-1/2}\right)^\top S(t)\left(\theta + h\overline{\varphi}(t)^{-1/2}\right) + \frac{1}{2}\theta^\top S(t)\theta \\ &= h^\top \overline{\varphi}(t)^{-1/2} A(t) - \theta^\top S(t) h^\top \overline{\varphi}(t)^{-1/2} \\ &- \frac{1}{2} h^\top \overline{\varphi}(t)^{-1/2} S(t) h \overline{\varphi}(t)^{-1/2}, \end{aligned}$$

which can be rewritten as

$$l(\theta + \delta(t)h, X_t) - l(\theta, X_t) = h^\top \frac{1}{\sqrt{\overline{\varphi}(t)}} A(t) - \frac{1}{\sqrt{\overline{\varphi}(t)}}\theta^\top S(t) h - \frac{1}{2\overline{\varphi}(t)} h^\top S(t) h. \tag{4.22}$$

Using the decomposition (4.19), we can rewrite $A(t)$ as

$$A(t) = S(t)^\top \theta + M(t), \tag{4.23}$$

where $M(t)$ is the N-dimensional martingale defined in (4.16). Replacing (4.23) in (4.22) and by Remark 4.4.2, we obtain

$$l(\theta + h\overline{\varphi}(t)^{-1/2}, X_t) - l(\theta, X_t) = \frac{1}{\overline{\varphi}(t)^{1/2}} h^\top M(t) - \frac{1}{2\overline{\varphi}(t)} h^\top [M](t) h. \tag{4.24}$$

We conclude from (4.24) and Lemma 4.4.1 that (1.2) holds, which implies that the LAN property is satisfied for all $\theta \in \operatorname{int} \Theta_0$.

■

Proof of Theorem 4.3.2

The step-by-step to proof the Theorem 4.3.2 is based on the proof of Mai [43, Theorem 3.5].

Before we prove the first technical lemma, we will recall the (upper bound of) well-known Mills ratio (see, for example, Feller [11, Lemma 2, p.175]). Let Z be a normal random variable under the probability measure P, with density $f_Z(\cdot)$ and set $z > 0$. Then

$$P\left(Z > z\right) < \frac{1}{z} f_Z(z). \tag{4.25}$$

Lemma 4.4.3. *Suppose that Assumption 4.2.1 holds. Then, for any $\beta \in (0, 1/2)$, we have*

$$P_{\theta_0}\left(|\Delta_i W + \Delta_i D| > \Delta_n^\beta\right) = O(\Delta_n^{2-2\beta}), \quad \text{as } \; n \to \infty,$$

for $i = 1, \cdots, n-1$, where $\mathbf{W}$ denotes the Gaussian component of $\mathbf{L}$ and $\mathbf{D}$ is the drift part of $\mathbf{X}$ defined in (4.13).

Proof. We will assume, without losing of generality, that $\sigma = 1$. Otherwise, the same approach could be developed giving the fact that $\sigma W(t) \sim N(0, \sigma^2 t)$. Observe that

$$\omega \notin \left[|\Delta_i W| > \frac{\Delta_n^\beta}{2}\right] \bigcup \left[|\Delta_i D| > \frac{\Delta_n^\beta}{2}\right]$$

implies

$$\omega \in \left[|\Delta_i W| \leq \frac{\Delta_n^\beta}{2}\right] \bigcap \left[|\Delta_i D| \leq \frac{\Delta_n^\beta}{2}\right]$$

and consequently

$$|\Delta_i W(\omega) + \Delta_i D(\omega)| \leq |\Delta_i W(\omega)| + |\Delta_i D(\omega)| \leq \Delta_n^\beta.$$

Thus,

$$P_{\theta_0}\left(|\Delta_i W + \Delta_i D| > \Delta_n^\beta\right) \leq P_{\theta_0}\left(|\Delta_i W| > \frac{\Delta_n^\beta}{2}\right) + P_{\theta_0}\left(|\Delta_i D| > \frac{\Delta_n^\beta}{2}\right).$$

By $\Delta_i W \sim N(0, \Delta_i)$, the symmetry of the normal distribution and the Mills ratio (4.25), it follows that

$$\begin{aligned} P_{\theta_0}\left(|\Delta_i W| \geq \frac{\Delta_n^\beta}{2}\right) &= 2P_{\theta_0}\left(\frac{1}{\sqrt{\Delta_i}}\Delta_i W \geq \frac{\Delta_n^\beta}{2\sqrt{\Delta_i}}\right) \\ &\leq 2\frac{2\sqrt{\Delta_i}}{\sqrt{2\pi}\Delta_n^\beta} \exp\left\{-\frac{1}{2}\left(\frac{\Delta_n^\beta}{2\sqrt{\Delta_i}}\right)^2\right\}. \end{aligned}$$

Since $\frac{2}{\sqrt{2\pi}}\frac{\Delta_i^{1/2}}{\Delta_n^\beta} \leq \Delta_n^{1/2-\beta}$ and $\frac{1}{2}\left(\frac{\Delta_n^\beta}{2\Delta_i^{1/2}}\right)^2 \geq \frac{1}{8\Delta_n^{1-2\beta}}$, we have

$$P_{\theta_0}\left(|\Delta_i W| \geq \frac{\Delta_n^\beta}{2}\right) \leq 2\Delta_n^{1/2-\beta} \exp\left\{-\frac{1}{8\Delta_n^{1-2\beta}}\right\}.$$

We claim that

$$P_{\theta_0}\left(|\Delta_i D| > \frac{\Delta_n^\beta}{2}\right) = O(\Delta_n^{2-2\beta}).$$

Indeed, by Jensen's Inequality for finite measures,

$$\begin{aligned} \left(\int_{t_i}^{t_{i+1}} b(\theta_0, X_s)ds\right)^2 &= \Delta_i^2 \left(\frac{1}{\Delta_i}\int_{t_i}^{t_{i+1}} b(\theta_0, X_s)ds\right)^2 \\ &\leq \Delta_i \int_{t_i}^{t_{i+1}} b^2(\theta_0, X_s)ds \\ &\leq \Delta_n \int_{t_i}^{t_{i+1}} b^2(\theta_0, X_s)ds, \end{aligned}$$

which, from (4.15), implies

$$\begin{aligned} E_{\theta_0}\left[(\Delta_i D)^2\right] &\leq \Delta_n E_{\theta_0}\left[\sum_{j=1}^{N}\sum_{k=1}^{N}\theta_j^0\theta_k^0\int_{t_i}^{t_{i+1}} b_j(X_s)b_k(X_s)ds\right] \\ &= \Delta_n \sum_{j=1}^{N}\sum_{k=1}^{N}\theta_j^0\theta_k^0\int_{t_i}^{t_{i+1}} E_{\theta_0}\left[b_j(X_s)b_k(X_s)\right]ds \\ &\leq \Delta_n \sum_{j=1}^{N}\sum_{k=1}^{N}\theta_j^0\theta_k^0\int_{t_i}^{t_{i+1}} E_{\theta_0}\left[b_j(X_s)b_k(X_s)\right]ds \\ &\leq \Delta_n^2 \sum_{j=1}^{N}\sum_{k=1}^{N}\theta_j^0\theta_k^0 \sup_{s\in[t_i,t_{i+1}]} E_{\theta}\left[b_j(X_s)b_k(X_s)\right]\cdot \end{aligned}$$

It follows from Markov Inequality

$$\begin{aligned} P_{\theta_0}\left(|\Delta_i D| > \frac{\Delta_n^{\beta}}{2}\right) &\leq \frac{E_{\theta_0}\left[(\Delta_i D)^2\right]}{\left(\frac{\Delta_n^{\beta}}{2}\right)^2} \\ &\leq 4\frac{\Delta_n^2}{\Delta_n^{2\beta}}\sum_{j=1}^{N}\sum_{k=1}^{N}\theta_j^0\theta_k^0 \sup_{s\in[t_i,t_{i+1}]} E_{\theta_0}\left[b_j(X_s)b_k(X_s)\right] \\ &= O(\Delta_n^{2-2\beta}), \end{aligned}$$

as claimed, hence proving the lemma.

■

Remark 4.4.4. *We assume Assumption 4.2.1 in Lemma 4.4.3. However, we do not need that decomposition explicitly. It suffices to assume that $b(\theta, X_s) \in L^2(dP_\theta \times ds)$.*

Define, for each $i \in \{1, \cdots, n\}$,

$$E_i := \left[\mathbf{1}_{\left[|\Delta_i X| \leq \Delta_n^{\beta}\right]} = \mathbf{1}_{[\Delta_i N = 0]}\right].$$

Lemma 4.4.5. *In addition to the hypothesis of Lemma 4.4.3, assume that $n\Delta_n t_n^{-1} = O(1)$, and $t_n \Delta_n^{1-2\beta} = o(1)$, as $n \to \infty$. Then*

$$P_{\theta_0}\left(\bigcap_{i=1}^{n} E_i\right) \longrightarrow 1 \ \ as \ \ n \to \infty.$$

Proof. Observe that

$$P_{\theta_0}\left(\left(\bigcap_{i=1}^{n} E_i\right)^c\right) = P_{\theta_0}\left(\bigcup_{i=1}^{n} E_i^c\right) \leq \sum_{i=1}^{n} P_{\theta_0}\left(E_i^c\right).$$

Taking $K_i = \left[|\Delta_i X| \leq \Delta_n^{\beta}\right]$ and $M_i = [\Delta_i N = 0]$, we can write E_i^c as

$$E_i^c = [\mathbf{1}_{K_i} \neq \mathbf{1}_{M_i}] = (K_i - M_i) \cup (M_i - K_i)\,. \tag{4.26}$$

Note that

$$\begin{aligned}
P_{\theta_0}\left(K_i - M_i\right) &= P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta}, \Delta_i N > 0\right) \\
&= \sum_{j=1}^{\infty} e^{-\lambda \Delta_i} \frac{(\lambda \Delta_i)^j}{j!} P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta} \middle| \Delta_i N = j\right) \\
&\le P_{\theta_0}\left(\Delta_i N = 1\right) P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta} \middle| \Delta_i N = 1\right) + O\left(\Delta_n^2\right).
\end{aligned}$$

We still have to

$$\begin{aligned}
P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta} \middle| \Delta_i N = 1\right) &= P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta}, |\Delta_i J| > 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right) \\
&+ P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta}, |\Delta_i J| \le 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right). \qquad (4.27)
\end{aligned}$$

We claim that the first term on the right side of (4.27) is bounded by $P_{\theta_0}\left(\Delta_i N = 1\right)^{-1} O(\Delta_n^{2-2\beta})$. Indeed, denote $p_i := P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta}, |\Delta_i J| > 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right)$, by triangular inequality,

$$\begin{aligned}
p_i &= P_{\theta_0}\left(|\Delta_i W + \Delta_i J + \Delta_i D| \le \Delta_n^{\beta}, |\Delta_i J| > 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right) \\
&= \frac{P_{\theta_0}\left(|\Delta_i W + \Delta_i J + \Delta_i D| \le \Delta_n^{\beta}, |\Delta_i J| > 2\Delta_n^{\beta}, \Delta_i N = 1\right)}{P_{\theta_0}\left(\Delta_i N = 1\right)} \\
&\le \frac{P_{\theta_0}\left(|\Delta_i W + \Delta_i D| > \Delta_n^{\beta}, \Delta_i N = 1\right)}{P_{\theta_0}\left(\Delta_i N = 1\right)},
\end{aligned}$$

hence, by Lemma 4.4.3,

$$\begin{aligned}
P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta}, |\Delta_i J| > 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right) &\le \frac{P_{\theta_0}\left(|\Delta_i W + \Delta_i D| > \Delta_n^{\beta}\right)}{P_{\theta_0}\left(\Delta_i N = 1\right)} \\
&= P_{\theta_0}\left(\Delta_i N = 1\right)^{-1} O(\Delta_n^{2-2\beta}).
\end{aligned}$$

Since F is the distribution of the size of the jumps of $\mathbf{J}$, we obtain that the second term in the right side of (4.27) satisfies

$$\begin{aligned}
P_{\theta_0}\left(|\Delta_i X| \le \Delta_n^{\beta}, |\Delta_i J| \le 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right) &\le P_{\theta_0}\left(|\Delta_i J| \le 2\Delta_n^{\beta} \middle| \Delta_i N = 1\right) \\
&= P_F\left(\left(-2\Delta_n^{\beta}, 2\Delta_n^{\beta}\right)\right)
\end{aligned}$$

hence,

$$\begin{aligned}
P_{\theta_0}\left(K_i - M_i\right) &\le O(\Delta_n^{2-2\beta}) + \Delta_n P_F\left(\left(-2\Delta_n^{\beta}, 2\Delta_n^{\beta}\right)\right) + O\left(\Delta_n^2\right) \\
&= O(\Delta_n^{2-2\beta}) + \Delta_n P_F\left(\left(-2\Delta_n^{\beta}, 2\Delta_n^{\beta}\right)\right). \qquad (4.28)
\end{aligned}$$

On the other hand, by Lemma 4.4.3,

$$\begin{aligned}
P_{\theta_0}\left(M_i - K_i\right) &= P_{\theta_0}\left(|\Delta_i X| > \Delta_n^{\beta}, \Delta_i N = 0\right) \\
&\le P_{\theta_0}\left(|\Delta_i W + \Delta_i D| > \Delta_n^{\beta}\right) \\
&= O(\Delta_n^{2-2\beta}),
\end{aligned}$$

which, together with (4.26) and (4.28), implies

$$P_{\theta_0}\left(E_i^c\right) \le O(\Delta_n^{2-2\beta}) + \Delta_n P_F\left(\left(-2\Delta_n^{\beta}, 2\Delta_n^{\beta}\right)\right)\cdot$$

Therefore, since $n\Delta_n t_n^{-1} = O(1)$, we have

$$P_{\theta_0}\left(\left(\bigcap_{i=1}^{n} E_i\right)^c\right) \le \sum_{i=1}^{n} P_{\theta_0}\left(E_i^c\right) \le O(t_n)P_F\left(\left(-2\Delta_n^{\beta}, 2\Delta_n^{\beta}\right)\right) + O(t_n\Delta_n^{1-2\beta}), \text{ as } n\to\infty,$$

which proves the lemma, provided that $P_F\left(-2\Delta_n^{\beta}, 2\Delta_n^{\beta}\right) = o(t_n^{-1})$ and $t_n\Delta_n^{1-2\beta} = o(1)$.

■

Lemma 4.4.6. *In addition to the hypothesis of Lemma 4.4.5, assume that the Lévy process* $\mathbf{L}$ *has characteristic triplet* $(0,\sigma^2,\nu)$ *and that* $\Delta_i W$*,* $\Delta_i D$ *and* $b_j(X_i)$ *are mutually independent, under* P_{θ_0}*, for each* $i\in\{1,\cdots,n-1\}$ *and* $j\in\{1,\cdots,N\}$*. Then, for each* $j\in\{1,\cdots,N\}$*,*

$$\left|\sum_{i=1}^{n-1} b_j(X_i)\left(\Delta_i X\mathbf{1}_{\left[|\Delta_i X|\le\Delta_n^{\beta}\right]} - \Delta_i X^c\right)\right| = O\left(t_n\Delta_n^{1/2}\right), \quad as\ t\to\infty\cdot$$

Proof. On $\bigcap_{i=1}^{n} E_i$ holds

$$\sum_{i=1}^{n-1} b_j(X_i)\left(\Delta_i X\mathbf{1}_{\left[|\Delta_i X|\le\Delta_n^{\beta}\right]} - \Delta_i X^c\right) = \sum_{i=1}^{n-1} b_j(X_i)\left(\Delta_i X\mathbf{1}_{[\Delta_i N=0]} - \Delta_i X^c\right)\cdot$$

Set $C_i = [\Delta_i N > 0]$. Then

$$\Delta_i X\mathbf{1}_{[\Delta_i N=0]} - \Delta_i X^c = -\Delta_i X^c\mathbf{1}_{C_i},$$

which implies

$$E_{\theta_0}\left|\mathbf{1}_{\bigcap_{i=1}^n E_i}\sum_{i=1}^{n-1} b_j(X_i)\left(\Delta_i X\mathbf{1}_{[\Delta_i N=0]} - \Delta_i X^c\right)\right| = E_{\theta_0}\left|\sum_{i=1}^{n-1} b_j(X_i)\Delta_i X^c\mathbf{1}_{C_i\cap\left(\bigcap_{i=1}^n E_i\right)}\right|$$

and, by triangular inequality and the decomposition $X^c = W + D$, we obtain

$$\begin{aligned} E_{\theta_0}\left|\sum_{i=1}^{n-1} b_j(X_i)\Delta_i X^c\mathbf{1}_{C_i\cap\left(\bigcap_{i=1}^n E_i\right)}\right| &\le \sum_{i=1}^{n-1} E_{\theta_0}\left[\left(|b_j(X_i)\Delta_i W + b_j(X_i)\Delta_i D|\right)\mathbf{1}_{C_i\cap\left(\bigcap_{i=1}^n E_i\right)}\right] \\ &\le \sum_{i=1}^{n-1} E_{\theta_0}\left[\left(|b_j(X_i)\Delta_i W| + |b_j(X_i)\Delta_i D|\right)\mathbf{1}_{C_i}\right]\cdot \end{aligned}$$

It follows, from independence of $\Delta_i W$, $\Delta_i N$ and $b_j(X_i)$, that

$$\sum_{i=1}^{n-1} E_{\theta_0}\left|b_j(X_i)\Delta_i W\mathbf{1}_{C_i}\right| = \sum_{i=1}^{n-1} E_{\theta_0}\left|b_j(X_i)\right| E_{\theta_0}\left|\Delta_i W\right| P_{\theta_0}\left(C_i\right)$$

which implies, from $P_{\theta_0}(C_i) \le \lambda\Delta_n$,

$$\begin{aligned} \sum_{i=1}^{n-1} E_{\theta_0}\left|b_j(X_i)\Delta_i W\mathbf{1}_{C_i}\right| &\le \sum_{i=1}^{n-1} E_{\theta_0}\left|b_j(X_i)\right| E_{\theta}\left|\Delta_i W\right| P_{\theta_0}(C_i) \\ &\le \max_{i=1,\cdots,n-1}\left\{E_{\theta_0}\left|b_j(X_i)\right| E_{\theta_0}\left|\Delta_i W\right|\right\}(n-1)\lambda\Delta_n \qquad (4.29) \\ &\le K_0(\theta_0)(n-1)\Delta_n^{1/2} \\ &= O\left(t_n\Delta_n^{1/2}\right), \end{aligned}$$

where $K_0(\theta_0)$ is a constant and the last equality holds since $n\Delta_n t_n^{-1} = O(1)$. Analogously, by Hölder's Inequality,

$$\sum_{i=1}^{n-1} E_{\theta_0} |b_j(X_i)\Delta_i D \mathbf{1}_{C_i}| \le \sum_{i=1}^{n-1} \left(E_{\theta_0} \left|b_j^2(X_i)\left(\Delta_i D\right)^2\right|\right)^{1/2} P_{\theta_0}\left(C_i\right)^{1/2} = O\left(t_n \Delta_n^{1/2}\right), \text{ as } n \to \infty\cdot$$

■

Consider the natural discretized estimator obtained from $\hat{\theta}(t)$, that is,

$$\bar{\theta}_n = S_n^{-1}\bar{A}_n,$$

where

$$\bar{A}_n = \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i X\right)_{j=1,\cdots,N}$$

and S_n is as in (4.12).

Lemma 4.4.7. *In addition to the hypothesis of Theorem 4.2.4, assume that Assumption 4.3.1 holds. Then*

$$\textit{Law}\left(t_n^{1/2}\left(\bar{\theta}_n - \theta_0\right)\middle| P_{\theta_0}\right) \to N(0, \Sigma(\theta_0)^{-1}), \ \textit{as} \ t \to \infty,$$

where $\Sigma(\theta_0)$ *is defined in Assumption 4.2.3.*

Proof. Note that, by (4.18), the continuous martingale part can be decomposed as

$$X^c(t) = D(t) + \sigma W(t),$$

where $\mathbf{W}$ is a Wiener process and $\mathbf{D}$ is defined in (4.13). Thus,

$$\frac{1}{\sigma^2}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i X^c = \frac{1}{\sigma^2}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i D + \frac{1}{\sigma}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i W\cdot$$

Therefore,

$$\begin{aligned} t_n^{1/2}\left(\bar{\theta}_n - \theta_0\right) &= t_n^{1/2}\left(S_n^{-1}\bar{A}_n - \theta_0\right) \\ &= t_n^{1/2}\left(S_n^{-1}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i D\right)_{N\times 1} - \theta_0\right) \\ &+ t_n^{1/2} S_n^{-1}\left(\frac{1}{\sigma}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i W\right)_{N\times 1}\cdot \end{aligned}$$

Observe that, for each $j = 1, \cdots, N$,

$$\frac{1}{t_n^{1/2}}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i W \sim \frac{1}{t_n^{1/2}}\int_0^{t_n} b_j(X_s)dW(s), \text{ as } n \to \infty,$$

which, together with $t_n S_n^{-1} \to \Sigma(\theta_0)^{-1}$, implies

$$\text{Law}\left(t_n S_n^{-1}\frac{1}{t_n^{1/2}}\left(\frac{1}{\sigma}\sum_{i=1}^{n-1} b_j(X_i)\Delta_i W\right)_{N\times 1}\middle| P_{\theta_0}\right) \to N\left(0, \Sigma(\theta_0)^{-1}\right), \text{ as } n \to \infty\cdot$$

It remains to be proved that

$$t_n^{1/2}\left(S_n^{-1}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\Delta_i D\right)_{N\times 1}-\theta_0\right)\to 0$$

in P_{θ_0}-probability as $n\to\infty$. Indeed, let us define a $N\times N$-matrix $\bar{S}_n$ that is asymptotically equivalent to S_n by

$$\bar{S}_n := \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\int_{t_i}^{t_{i+1}}b_k(X_s)ds\right)_{N\times N}.$$

Then,

$$\begin{aligned}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\Delta_i D\right)_{N\times 1} &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\int_{t_i}^{ti+1}b(\theta_0,X_s)ds\right)_{N\times 1}\\ &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\int_{t_i}^{ti+1}b(\theta_0,X_s)ds\right)_{N\times 1},\end{aligned}$$

by Assumption 4.2.1 we have

$$\begin{aligned}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\Delta_i D\right)_{N\times 1} &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\sum_{k=1}^{N}\theta_k^0\int_{t_i}^{ti+1}b_k(X_s)ds\right)_{N\times 1}\\ &= \left(\sum_{k=1}^{N}\theta_k^0\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\int_{t_i}^{ti+1}b_k(X_s)ds\right)_{N\times 1}\\ &= \left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\int_{t_i}^{ti+1}b_k(X_s)ds\right)_{N\times 1}\theta_0\\ &= \bar{S}_n\theta_0,\end{aligned}$$

which implies

$$\begin{aligned}t_n^{1/2}\left(S_n^{-1}\left(\frac{1}{\sigma^2}\sum_{i=1}^{n-1}b_j(X_i)\Delta_i D\right)_{N\times 1}-\theta_0\right) &= t_n S_n^{-1}\frac{1}{t_n^{1/2}}\bar{S}_n\theta_0 - t_n S_n^{-1}\frac{1}{t_n^{1/2}}S_n\theta_0\\ &= t_n S_n^{-1}\frac{1}{t_n^{1/2}}\left(\bar{S}_n - S_n\right)\theta_0.\end{aligned}$$

The statement follows since $\frac{1}{t_n^{1/2}}\left(\bar{S}_n - S_n\right)\to 0$ in P_{θ_0}-probability as $n\to\infty$.

■

Now we are able to prove Theorem 4.3.2.

Proof of Theorem 4.3.2. Observe that

$$t_n^{1/2}\left(\hat{\theta}_n^{FMLE}-\theta_0\right) = t_n^{1/2}\left(\hat{\theta}_n^{FMLE}-\bar{\theta}_n\right) - t_n^{1/2}\left(\hat{\theta}_n^{FMLE}-\theta_0\right).$$

By Lemma 4.4.7,

$$\text{Law}\left(t_n^{1/2}\left(\bar{\theta}_n-\theta_0\right)\Big| P_{\theta_0}\right)\to N(0,\Sigma(\theta_0)^{-1}),\quad \text{as } t\to\infty.$$

Finally, by Lemma 4.4.6, we have

$$\begin{aligned} t_n^{1/2}\left(\hat{\theta}_n^{FMLE} - \bar{\theta}_n\right) &= t_n^{1/2}\left(S_n^{-1}A_n - S_n^{-1}\bar{A}_n\right) \\ &= t_n S_n \frac{1}{t_n}\left(A_n - \bar{A}_n\right) \\ &= o_{P_{\theta_0}}(1), \end{aligned}$$

which proves the theorem.

■

Chapter 5

Conclusion

Our studies aimed to investigate the estimation of the drift parameter of GLE with Lévy jumps, which is non-Markovian and non-Gaussian continuous time stochastic process. The results indicated that the MLE for the GLE is consistent, asymptotically normal and efficient. Further findings showed that this estimator, under filtering "big" jumps, can be applied to estimate drift parameter from an associated discrete time process.

Research focused on statistical inference issues for the GLE have not been widely explored in the literature yet. We proposed estimation methods for several GLE models, without initially requiring knowledge of the explicit GLE solution.

We analyzed the influence of jumps in the estimation through simulations and concluded that a higher frequency of jumps did not interfere in the estimation of the 3-parameter GOU-FE. On the other hand, certain regions of the parameter space have better estimates than others. This was expected since parameter space restrictions are necessary to guarantee ergodicity and good asymptotic properties for the GLE.

We summarize some of our main results bellow:

1. We extended the studies on estimators for the Langevin equation and SDDE. More precisely, we studied MLE for the drift parameter of the GLE observed continuously in time. In Theorem 2.3.4 we showed that with appropriated convergence assumptions we have a consistent and asymptotically normal MLE which is also efficient in the sense of Hájek-Le Cam convolution theorem.

2. In Section 2.4, a filtered MLE (FMLE) was proposed for cases where the GLE is observed on discrete times. We evaluated the results of this discretization via studies of simulations of the GOU-FE process. Potentials and limitations (computational time and optimization difficulties) of the applications of these models were presented.

3. In Section 3.2, a new process was proposed generalizing the GOU-FE process (called *3-parameter GOU-FE*). In Theorem 3.2.1 we showed that this process is a solution of a class of GLE. Furthermore, an order 3 autoregressive form for these processes were obtained in Proposition 3.2.2. An autoregressive form and its error distribution for the GOU process in its general form was also addressed (Theorem 3.2.6 and Proposition 3.2.7).

4. In Section 3.3, a modified FMLE (*mFMLE*) was proposed by introducing the information of the simulated Lévy path. The FMLE and mFMLE have so far been studied only through simulations of

the 3-parameter GOU-FE. Our simulations showed that mFMLE can indeed improve the estimation in relation to FMLE, however a previous study of Lévy noise path is necessary.

5. The goal of Chapter 4 was to study the theoretical asymptotic behavior of the FMLE. For this, based on Langevin equation, Cosine process and SDDE, we considered a particular class of GLE for which the drift parameter can be "separated" of the process information (Assumption 4.2.1). Then, an explicit MLE was found and Theorem 4.2.4 presented its asymptotic behaviour, which is the same of Theorem 2.3.4 but with less assumptions and technical difficulties. In Theorem 4.3.2 we showed that, with a suitable discrete time information, the FMLE inherits the asymptotic properties of the MLE. Unfortunately, despite the technical simplicity, this result cannot be applied to the GOU-FE process.

Based on these conclusions, we finish this chapter presenting a list of open problems related to the results obtained in previous chapters.

i. **Ergodicity** (in the sense of Birkhoff's theorem) proved to be an essential condition for studying the asymptotic behavior of the MLE. What can we say about ergodicity for GLE with Lévy jumps? How can we verify this property for a real temporal series?

ii. **Convergence rates** are natural questions for the asymptotic behaviour of the MLE since a version of CLT was proved (Theorem 2.3.4). Can we obtain a Large Deviation theorem for the MLE?

iii. In Proposition 3.2.2 the autoregressive form of the discrete time process does not have a white noise error. Alcântara [1] proposed use the model $ARMA(p, p-1)$ for describe better the GOU-FE with two parameters. How can we approximate the **error distribution** of the autoregressive form?

iv. Knowing a good **noise approximation** proved to be important for the mFMLE estimator. Thus, this becomes a fundamental initial step for the estimator. What is the best way to simulate the Lévy path for real applications?

v. The memory function found in Theorem 3.2.1 has a very similar form of the covariance function of the stationary solution from a SDDE (see [30, Equation (9)]). What are the **connections between the memory function of the GOU-FE and the autocorrelation function of the SDDE**?

vi. **Hypotheses tests** for the drift parameter estimation is an open problem.

vii. It is possible to use a **Bayesian approach** for estimating the drift parameter of the GLE?

Appendix A

Review of the Theory of Stochastic Processes and Statistical Inference

In this appendix we present a revision of the main theoretical results used in this thesis, with the purpose that this work be self-contained. As these are very technical topics, we suggest a detailed reading of the following references Applebaum [2], Ibragimov and Has'minskii [22], Jacod and Shiryaev [23], Klebaner [27], Küchler and Sϕrensen [33], Le Cam and Yang[36] and Vaart[65]. Other references are suggested throughout this appendix to consult some specific results.

A.1 Basic Definitions and Properties

Consider a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$.

Definition A.1.1 (Usual Conditions)**.** *We say that a filtration $\{\mathcal{F}_t; t \geq 0\}$ satisfies the usual conditions if*

1. *it is right-continuous, i.e., $\mathcal{F}_t = \mathcal{F}_{t+} = \bigcup_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$;*
2. *$\mathcal{F}_0$ contains all the $\mathbb{P}$-negligible ($\mathbb{P}$-null) events in $\mathcal{F}$, i.e., if $N \in \mathcal{F}$ is such that $\mathbb{P}(N) = 0$, then $N \in \mathcal{F}_0$.*

A *random set* A (i.e. a subset of $\Omega \times \mathbb{R}_+$) is called *evanescent* if the set

$$\{\omega; \exists t \in \mathbb{R}_+ \text{ with } (\omega, t) \in A\}$$

is $\mathbb{P}$-null.

Consider two processes $\mathbf{X} = \{X(t); t \geq 0\}$ and $\mathbf{Y} = \{Y(t); t \geq 0\}$. We say that $\mathbf{Y}$ is a *modification* of $\mathbf{X}$ if, for every $t \geq 0$, we have

$$P\left(X(t) = Y(t)\right) = 1.$$

X and Y are *indistinguishable* if

$$P\left(X(t) = Y(t); 0 \leq t < \infty\right) = 1.$$

Consider a family of measures $\{P_\theta; \theta \in \Theta\}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ with parameter set $\Theta \subset \mathbb{R}^N$ such that the interior int Θ is non-empty. We assume that the filtration $\{\mathcal{F}_t; t \geq 0\}$ satisfies the usual conditions.

Definition A.1.2. *We say that P_θ is* locally absolutely continuous *with respect to P_{θ_0} and write $P_\theta \overset{loc}{\ll} P_{\theta_0}$ if $P^t_\theta << P^t_{\theta_0}$ for all $t \in \mathbb{R}_+$, where $P^t_\theta := P_\theta|_{\mathcal{F}_t}$ denotes the restriction of P_θ to $\mathcal{F}_t$.*

Usually, a "local" property is localized along a sequence of stopping times. The notion $\overset{loc}{\ll}$ indeed satisfies the same rule.

Lemma A.1.3 (Lemma III.3.3 in Jacod and Shiryaev [23])**.** *$P_\theta \overset{loc}{\ll} P_{\theta_0}$ if and only if there is an increasing sequence $\{T_n; n \in \mathbb{N}\}$ of stopping times, such that $T_n \uparrow \infty$ P_θ-a.s. and $P^{T_n}_\theta \ll P^{T_n}_{\theta_0}$ for all $n \in \mathbb{N}$.*

Definition A.1.4. *We say that a process $\mathbf{X} = \{X(t); t \geq 0\}$ is a* càd *(resp.* càg*, resp.* càdlàg*)* process *if all its paths are right-continuous (resp. are left-continuous, resp. are right-continuous and admit left-hand limits).*

Definition A.1.5. *We say that a filtration $\{\mathcal{F}_t; t \geq 0\}$ is* generated *by a process $\mathbf{X} = \{X(t); t \geq 0\}$ and a σ-field $\mathcal{G}$ if*

1. *$\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}^0_s$ and $\mathcal{F}^0_s = \mathcal{G} \vee \sigma(X(r); r \leq s)$ (in other words, $\{\mathcal{F}_t; t \geq 0\}$ is the smallest filtration such that $\mathbf{X}$ is adapted and $\mathcal{G} \subset \mathcal{F}_0$);*

2. *$\mathcal{F} = \mathcal{F}_{\infty^-}$ $(= \bigvee_t \mathcal{F}_t)$.*

Definition A.1.6.
1. *We call* optional σ-field *a σ-field $\mathcal{O}$ on $\Omega \times \mathbb{R}_+$ that is generated by all càdlàg adapted process.*

2. *We say that a* σ-field *a σ-field $\mathcal{P}$ on $\Omega \times \mathbb{R}_+$ is a* predictable σ-field *if it is generated by all càd adapted process.*

3. *We put $\tilde{\Omega} = \Omega \times \mathbb{R}_+ \times E$ with the σ-field $\tilde{\mathcal{O}} = \mathcal{O} \times \mathbb{B}(E)$ and $\tilde{\mathcal{P}} = \mathcal{P} \times \mathbb{B}(E)$. A function Y on $\tilde{\Omega}$ that is $\tilde{\mathcal{O}}$-measurable (resp.$\tilde{\mathcal{P}}$-measurable) is called an* optional *(resp. a* predictable*) function.*

We denote by $\mathcal{V}$ the class of all real-valued process $\mathbf{A} = \{A(t); t \geq 0\}$ that are càdlàg, adapted, with $A(0) = 0$ and whose each path has finite variation over each interval $[0, t]$.

Let $\mathbf{A} \in \mathcal{V}$. For each $\omega \in \Omega$, the path $t \mapsto A(\omega, t)$ is the distribution function of a signed measure (a positive measure if $\mathbf{A}$ is increasing) on $\mathbb{R}$ that is finite on each interval $[0, t]$. Denote this measure by $dA(\omega, t)$.

Let $\mathbf{A} \in \mathcal{V}$ and let $\mathbf{H} = \{H(t); t \geq 0\}$ be an optional process. By Jacod and Shiryaev [23] I.1.21, $t \mapsto H(\omega, t)$ is Borel measurable. Thus, we can define the *integral process*, denoted by $H \cdot A$, as follows

$$H \cdot A(\omega, t) := \begin{cases} \displaystyle\int_0^t H(\omega, s) dA(\omega, s) & if \quad \displaystyle\int_0^t |H(\omega, s)|\, dA(\omega, s) < \infty, \\ +\infty & \text{othewise.} \end{cases} \tag{A.1}$$

Proposition A.1.7 (Proposition I.3.5 in Jacod and Shiryaev [23])**.** *Consider $\mathbf{A} \in \mathcal{V}$ and let $\mathbf{H} = \{H(t); t \geq 0\}$ be an optional process, such that $\mathbf{B} = \mathbf{H} \cdot \mathbf{A}$ is finite-value. Then $\mathbf{B} \in \mathcal{V}$ and $dB \ll dA$. Moreover, if $\mathbf{A}$ and $\mathbf{H}$ are predictable, then $\mathbf{B}$ is also predictable.*

Let $\mathcal{A}_{loc}$ the class of all process with locally integrable variation.

Proposition A.1.8 (Theorem I.3.18 in Jacod and Shiryaev [23]). *If $\mathbf{A} \in \mathcal{A}_{loc}$, then there exists a process, called* compensator of $\mathbf{A}$ *and denoted by $\mathbf{A}^p$, which is unique up to an evanescent set, and which is characterized by being a predictable process of $\mathcal{A}_{loc}$ that $\mathbf{A} - \mathbf{A}^p$ is a local martingale.*

Consider $\mathbb{B}(\mathbb{R}^d)$ the Borel σ-field of $\mathbb{R}^d$. For a subset $E \subset \mathbb{R}^d$, we will denote by $\mathbb{B}(E)$ a σ-field of subsets of E (but not necessarily the Borel σ-field).

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ be a filtered probability space and consider $(E, \mathbb{B}(E))$ be a Blackwell space (here it is enough to know that $(\mathbb{R}^d, \mathbb{B}(\mathbb{R}^d))$ is a Blackwell space).

Definition A.1.9. *A* random measure *on $\mathbb{R}_+ \times E$ is a family*

$$\mu = \{\mu(\omega; dt, dx); \omega \in \Omega\}$$

of non-negative measures on $\left(\mathbb{R}_+ \times E, \mathbb{B}(\mathbb{R}_+) \otimes \mathbb{B}(E)\right)$ satisfying $\mu(\omega; \{0\} \times E) = 0$ identically.

Definition A.1.10. *An* integral-valued random measure *is a random measure that satisfies*

1. *$\mu(\omega; \{t\} \times E) \leq 1$ identically;*
2. *for each $A \in \mathbb{B}(\mathbb{R}_+) \otimes \mathbb{B}(E)$, $\mu(.; A)$ takes values in $\bar{\mathbb{N}}$;*
3. *μ is optional and $\tilde{\mathcal{P}}$-σ-finite.*

Proposition A.1.11 (Proposition II.1.16 in Jacod and Shiryaev [23]). *Let $\mathbf{X} = \{X(t); t \geq 0\}$ be an adapted càdlàg $\mathbb{R}^d$-valued process. Then*

$$\mu^X(\omega; dt, dx) = \sum_s \mathbf{1}_{[\Delta X(\omega,s)]} \varepsilon_{(s, \Delta X(\omega,s))}(dt, dx),$$

defines an integer valued random measure on $\mathbb{R}_+ \times \mathbb{R}^d$, where ε denotes the Dirac measure at point a.

Let ν denote the compensator of μ. By II.1.17 in Jacod and Shiryaev [23], there exists a version of ν such that $\nu(\omega; \{t\} \times E) \leq 1$. We define a predictable process $\mathbf{a} = \{a(t); t \geq 0\}$ by

$$a(\omega, t) := \nu(\omega; \{t\} \times E). \tag{A.2}$$

For each measurable function Y on $\tilde{\Omega}$ we define

$$\hat{Y}(\omega, t) := \begin{cases} \displaystyle\int_E Y(\omega, t, x) \nu(\omega; \{t\} \times dx) & \text{if this integral converges,} \\ +\infty & \text{otherwise.} \end{cases} \tag{A.3}$$

A.2 Semimartingale

Definition A.2.1. *An adapted process $\mathbf{M} = \{M(t); t \geq 0\}$ is called a* local martingale *if there exists a sequence of stopping times $\{T_n; n \in \mathbb{N}\}$ such that $T_n \uparrow \infty$ and for each n the stopped processes $M(t \wedge T_n)$ is a uniformly integrable martingale in t.*

Definition A.2.2. *We say that a process $\mathbf{X} = \{X(t); t \geq 0\}$ is a* semimartingale *if it is an adapted process such that, for each $t \geq 0$,*

$$X(t) = M(t) + V(t) \tag{A.4}$$

where $\mathbf{M} = \{M(t), t \geq 0\}$ is a local martingale and $\mathbf{V} = \{V(t); t \geq 0\}$ is an adapted process of finite variation.

Definition A.2.3. *We call a* truncation function *all* $h : \mathbb{R}^d \to \mathbb{R}^d$ *which are bounded and satisfy* $h(x) = x$ *in a neighborhood of* 0.

Definition A.2.4. *Let* h *be a truncation function. We call* characteristics (associated with h) *of a semimartingale* $\mathbf{X} = \{X(t); t \geq 0\}$ *the triplet* (B, C, ν) *consisting in:*

1. $B = B(h) = (B^i)_{i \leq d}$ *is a predictable random process of finite variation* $\mathbf{V}$ *appearing in (A.4);*
2. $C = (C^{ij})_{i,j \leq d}$ *given by*
$$C^{ij} = [X^{i,c}, X^{j,c}],$$
where X^c *is the continuous martingale part of* $\mathbf{X}$.
3. ν *is a predictable random measure on* $\mathbb{R}_+ \times \mathbb{R}^d$*, namely the compensator of the random measure* μ^X *associated to the jumps of* $\mathbf{X}$*, where*
$$\mu^X(\omega; dt \times dx) = \sum_s \mathbf{1}_{[\Delta X(\omega,s) \neq 0]} \varepsilon_{(s, \Delta X(\omega,s))}(dt, dx).$$

Consider a d-dimensional semimartingale $\mathbf{X} = (X_i)_{i \leq d}$ on $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, P_{\theta_0})$ with characteristics (B^0, C^0, ν^0) relative to a given truncation function h. Let $\mathbf{A} = \{A(t); t \geq 0\}$ be an increasing predictable process such that $C^0_{ij} = c_{ij} \cdot \mathbf{A}$.

Proposition A.2.5 (Girsanov's Theorem. See Theorem III.3.24 in Jacod and Shiryaev [23]). *Assume that* $P_\theta \overset{loc}{\ll} P_{\theta_0}$*, and let* $\mathbf{X} = \{X(t); t \geq 0\}$ *be as above. There exists a* $\tilde{\mathcal{P}}$*-measurable non-negative function* $Y(\cdot, \cdot)$ *and a predictable process* $\beta = (\beta^i)_{i \leq d}$ *satisfying*

$$|h(x)(Y-1)| * \nu^0_t < \infty \;\; P_\theta - a.s. \;\; for \;\; t \in \mathbb{R}_+,$$

$$\left| \sum_{j \leq d} c_{ij} \beta_j \right| \cdot A(t) < \infty \;\; and \;\; \left(\sum_{j,k \leq d} \beta_j c_{jk} \beta^k \right) \cdot A(t) < \infty \; P_\theta - a.s. \;\; for \;\; t \in \mathbb{R}_+,$$

and such that a version of the characteristics of $\mathbf{X}$ *relative to* P_θ *are*

$$\begin{cases} B_i &= B^0_i + \left(\sum_{j \leq d} c_{ij} \beta_j \right) \cdot A + h_i(x)(Y-1) * \nu^0, \\ C &= C^0, \\ \nu &= Y \cdot \nu^0. \end{cases} \tag{A.5}$$

Consider $\mathbf{X} = \{X(t); t \geq 0\}$ to be a process on a filtered space $\left(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P}\right)$. Let $\{\mathcal{G}_t; t \geq 0\}$ be a filtration generated by a process $\mathbf{X}$ and a sub-σ-field $\mathcal{H} \subset \mathcal{F}$. To emphasize the fact that $\{\mathcal{G}_t; t \geq 0\}$ is not right-continuous, in general, Jacod and Shiryaev [23] gave a specific name to the $\mathcal{G}_t$-stopping times.

Definition A.2.6. *A mapping* $T : \Omega \to \bar{\mathbb{R}}_+$ *such that* $[T \leq t] \in \mathcal{G}_t$ *for all* $t \in \mathbb{R}_+$ *is called a* $\mathcal{G}_t$-strict stopping time.

Assume that (c, A, β, Y) are given by Theorem A.2.5 and $\mathbf{a} = \{a(t); t \geq 0\}$ and $\hat{Y}(\cdot, \cdot)$ are given as in (A.2) and (A.3), respectively. Set

$$\tau = \inf \left\{ t; \text{ either } \left(\hat{Y}(t) > 1 \right) \text{ or } \left(a(t) = 1 \text{ and } \hat{Y}(t) < 1 \right) \right\}$$

a positive predictable time. Define

$$\mathbf{H} = (\beta c \beta)\mathbf{1}_{[0,\tau)} \cdot \mathbf{A} + \left(1-\sqrt{Y}\right)^2 \mathbf{1}_{[0,\tau)} * \nu + \sum_{s\leq \cdot} \left(\sqrt{1-a(s)} - \sqrt{1-\hat{Y}(s)}\right)^2 \mathbf{1}_{[s<\tau]}$$

and

$$T = \inf\{t; H(t) = \infty\}.$$

We have that $\mathbf{H} = \{H(t); t \geq 0\}$ is a predictable and *generalized increasing process*, i.e., it is a $\bar{\mathbb{R}}_+$-valued, $H(0) = 0$, its paths are non-decreasing and it is right-continuous on $[0, T)$ (and of course on (T, ∞)). We define a sequence of strict stopping times and a random set

$$T_n = \inf\{t; H(t) \geq n\}$$

and

$$\Delta = [0, \tau] \cap \left(\bigcup_n [0, T_n]\right).$$

Proposition A.2.7 (Proposition III.5.10 in Jacod and Shiryaev [23]). *Assume the above conditions. There is a process* $\mathbf{U} = \{U(t); t \geq 0\}$*, unique (up to* P*-indistinguishability) on the set* Δ*, such that for every stopping time* S *such that* $[0, S] \subset \Delta$*, the stopped process* $\mathbf{U}(S)$ *is the following* P_θ*-local martingale*

$$U(S) = (\beta \mathbf{1}_{[0,S]}) \cdot X^c + \left(Y - 1 + \frac{\hat{Y} - a}{1-a}\mathbf{1}_{[a<1]}\right)\mathbf{1}_{[0,S]} * (\mu - \nu).$$

Construction of the stochastic integral of locally bounded predictable processes with respect to a semimartingale can be find in Jacod and Shiryaev [23, Section I.4d] and Protter [54, Chapter II].

A.3 Lévy Process

In this section, we present some notions around the Lévy processes that are a particular case of semimartingales. For a more careful study of the Lévy processes, we recommend Applebaum [2], Sato [26].

Let $\mathcal{M}_1(\mathbb{R}^d)$ denote the set of all Borel probability measures on $\mathbb{R}^d$. We say that $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ is *infinitely divisible* if for any $n \in \mathbb{N}$ there exists $\mu_n \in \mathcal{M}_1(\mathbb{R}^d)$ such that $\mu = \mu_n^n := \underbrace{\mu * \cdots * \mu}_{n}$. Equivalently, μ is infinitely divisible if, and only if, for each $n \in \mathbb{N}$ there exists $\mu_n \in \mathcal{M}_1(\mathbb{R}^d)$ for which the characteristic functions satisfy

$$\phi_\mu(u) = [\phi_{\mu_n}(u)]^n$$

for each $u \in \mathbb{R}^d$.

Let ν be a Borel measure on $\mathbb{R}^d - \{0\}$. We say that it is a *Lévy measure* if

$$\int_{\mathbb{R}^d - \{0\}} (|y|^2 \wedge 1)\nu(dy) < \infty.$$

Theorem A.3.1 (Lévy-Khintchine, Theorem 1.2.14 in Applebaum [2]). $\mu \in \mathcal{M}_1(\mathbb{R}^d)$ *is infinitely divisible if there exists a vector* $b \in \mathbb{R}^d$*, a positive definite symmetric* $d \times d$ *matrix* A *and a Lévy measure* ν *on* $\mathbb{R}^d - \{0\}$ *such that, for all* $u \in \mathbb{R}^d$*,*

$$\phi_\mu(u) = \exp\left\{ i(b,u) - \frac{1}{2}(u, Au) + \int_{\mathbb{R}^d - \{0\}} \left[e^{i(u,y) - 1 - i(u,y)\xi_{B_1(0)}(y)}\right] \nu(dy) \right\}. \quad \text{(A.6)}$$

Conversely, any mapping of the form (A.6) is the characteristic function of an infinitely divisible probability measure on $\mathbb{R}^d$.

In Theorem A.3.1, $(\cdot,\cdot)$ denotes the inner product in $\mathbb{R}^d$. We call (b, A, ν) the *characteristic triplet* associated with μ.

Definition A.3.2. *Let* **L**$= \{L(t); t \geq 0\}$ *be a stochastic process defined on a probability space* $(\Omega, \mathcal{F}, \mathbb{P})$*. We say that* $\mathbf{L}$ *is a* **Lévy process** *if*

L1. $L(0) = 0$ *(a.s.);*

L2. **L** *has independent increments, i.e.,* $L(t_0), L(t_1) - L(t_0), \cdots L(t_n) - L(t_{n-1})$ *are independent random variables for every* $0 < t_0 < t_1 < \cdots < t_{n-1} < t_n$ *and for all positive integer* n*;*

L3. **L** *has stationary increments, i.e., for all* $t \geq 0$*,* $L(t+h) - L(t)$ *has the same distribution as* $L(h)$ *for all* $h > 0$*;*

L4. **L** *is stochastically continuous, i.e., for all* $\delta > 0$ *and* $s \geq 0$

$$\lim_{t \to s} \mathbb{P}\left(|L(t) - L(s)| > \delta\right) = 0;$$

L5. **L** *has càdlàg paths.*

Example A.3.3. *The following processes are examples of Lévy process:*

1. *a standard Brownian motion (or Wiener process)* **W** $= \{W(t); t \geq 0\}$

$$W(t) - W(s) \sim N(0, |t-s|);$$

2. *a Poisson process* **N** $= \{N(t); t \geq 0\}$ *of intensity* $\lambda > 0$

$$N(t) - N(s) \sim \textit{Poisson}(\lambda|t-s|);$$

3. *a Compound Poisson process*

$$\tilde{N}(t) = \sum_{j=1}^{N(t)} X_j$$

where $\{X_n; n \in \mathbb{N}\}$ *is a sequence of i.i.d. random variables;*

Proposition A.3.4 (Proposition 1.3.1 in Applebaum [2]). *If* $\mathbf{L}$ *is a Lévy process, then* $L(t)$ *is infinitely divisible for each* $t \geq 0$*.*

Proposition A.3.5 (Lévy-Itô's Decomposition, Theorem 2.4.16 in Applebaum [2]). *If* $\mathbf{X} = \{X(t); t \geq 0\}$ *is a Lévy process, then there exists* $b \in \mathbb{R}^d$*, a Brownian motion* $\{B_A(t); t \geq 0\}$ *with covariance matrix* A *and an independent Poisson random measure* N *on* $\mathbb{R}_+ \times (\mathbb{R}^d - \{0\})$ *such that, for each* $t \geq 0$*,*

$$X(t) = bt + B_A(t) + \int_{|x|<1} x\tilde{N}(t, dx) + \int_{|x|\geq 1} xN(t, dx)\cdot$$

Corollary A.3.6 (Corollary 2.4.21 in Applebaum [2]). *The Lévy characteristics* (b, A, ν) *of a Lévy process are uniquely determined by the process.*

Theorem A.3.7 (Theorem 2.4.25 in Applebaum [2]). *A Lévy process with characteristics* (b, A, ν) *has a finite variation if, and only if,* $A = 0$ *and* $\int_{|x|<1} |x|\nu(dx)\infty$*.*

Proposition A.3.8 (Proposition 2.7.1 in Applebaum [2]). *Every Lévy process is a semimartingale.*

A.4 Martingale Problem

Definition A.4.1. *Suppose that* $\mathbf{X} = \{X(t); t \geq 0\}$ *is a càdlàg function on* $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\})$. *A measure* $\mathbb{P}$ *on* $(\Omega, \mathcal{F})$ solves the martingale problem *associated with* $\mathbf{X}$*, an initial distribution* π *on* $(\Omega, \mathcal{F}_0)$ *and a triplet* (B, C, ν) *if*

1. *Under* $\mathbb{P}$ *the distribution of* $X(0)$ *equals* π*;*

2. $\mathbf{X}$ *is a semimartingale on* $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ *with characteristics* (B, C, ν) *relative to* h*.*

We denote by $s(\mathbf{X}|\pi; B, C, \nu)$ the set of all solution measures $\mathbb{P}$ of the martingale problem associated with the process $\mathbf{X}$, initial distribution π and characteristics (B, C, ν).

Definition A.4.2. *Let* $\{\mathcal{G}_t; t \geq 0\}$ *denote the filtration generated by* $\mathbf{X} = \{X(t); t \geq 0\}$*. We say that* local uniqueness *holds for a martingale problem* $s(\mathbf{X}|\pi; B, C, \nu)$ *if for every strict stopping time* T *any two solutions* $\mathbb{P}, \mathbb{P}' \in s(\mathbf{X}(T)|\pi; B_T, C_T, \nu_T)$ *of the stopped problem coincide on* $\mathcal{G}_T$*, where* $\mathbf{X}(T) = \{X(T \wedge t); t \geq 0\}$ *denotes the stopped process.*

Theorem A.4.3 (Theorem III.5.32 in Jacod and Shiryaev [23])**.** *Assume that* (B, C, ν) *and* (β, Y) *are given by Theorem A.2.5,* $\{\mathcal{F}_t; t \geq 0\}$ *is generated by* $\mathbf{X} = \{X(t); t \geq 0\}$*. Let* $\Delta = \bigcup_n [0, T_n]$ *up to* P_θ*-evanescent set where* $\{T_n; n \in \mathbb{N}\}$ *is a sequence of strict stopping times. Suppose that local uniqueness holds for the martingale problem* $s(\mathbf{X}|\pi; B, C, \nu)$*, with* P_θ *as its unique solution. If* $P_\theta \overset{loc}{\ll} P_{\theta_0}$*, then the* density process $\mathbf{Z} = \{Z(t); t \geq 0\}$ *of* P_θ *relative to* P_{θ_0} *is given by*

$$Z(t) = \begin{cases} Z_0 \exp\left\{U(t) - \frac{1}{2}(\beta c \beta) \cdot A(t)\right\} \prod_{s \leq t} (1 + \Delta U(s))\, e^{-\Delta U(s)}, & t \in \Delta, \\ 0, & t \notin \Delta. \end{cases}$$

where Z_0 *is (*P_{θ_0}*-a.s. equals) the Radon-Nikodym derivative* $Z_0 = d\pi/d\pi_0$*.*

A.5 Hellinger Processes and Absolute Continuity of Measures

We consider a filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\})$ with $\mathcal{F} = \mathcal{F}_{\infty^-}$ and two fixed probability measure P_θ and P_{θ_0} on $(\Omega, \mathcal{F})$. Instead of $P_\theta \ll \mu$ and $P_{\theta'} \ll \mu$, we will assume

$$P_\theta \overset{loc}{\ll} \mu \text{ and } P_{\theta_0} \overset{loc}{\ll} \mu.$$

We call $\mathbf{Z} = \{Z(t); t \geq 0\}$ and $\mathbf{Z}_0 = \{Z_0(t); t \geq 0\}$ the *density process* of P_θ and P_{θ_0}, relative to μ. By III.3.4 in Jacod and Shiryaev [23], they are μ-martingale.

Set

$$\begin{cases} R_n &= \inf\left\{t; Z(t) < \frac{1}{n}\right\}, \\ R'_n &= \inf\left\{t; Z_0(t) < \frac{1}{n}\right\}, \\ \Gamma'' &= \left(\bigcup_n [0, R_n]\right) \bigcap \left(\bigcup_n [0, R'_n]\right). \end{cases}$$

Theorem A.5.1 (Theorems IV.1.18 and IV.1.22 in Jacod and Shiryaev [23])**.** *Let* $\alpha \in (0,1)$ *and* $\mathbf{Y}(\alpha) = \mathbf{Z}^\alpha \mathbf{Z}_0^{1-\alpha}$*. There exists a predictable increasing* $\bar{\mathbb{R}}_+$*-valued process* $\mathbf{H}(\alpha; \theta, \theta_0) = \{H(\alpha; \theta, \theta_0, t); t \geq 0\}$

unique up to μ-indistinguishable, called Hellinger process of order α between P_θ and P_{θ_0}, which meets $H(\alpha;\theta,\theta_0,0)=0$ and the following two conditions

$$H(\alpha;\theta,\theta_0,t)=\mathbf{1}_{\Gamma''}\cdot H(\alpha;\theta,\theta_0,t)$$

and

$$Y(\alpha,t)+Y(\alpha,t^-)\cdot H(\alpha;\theta,\theta_0,t)$$

is a μ-martingale. Moreover, $\mathbf{H}(\alpha;\theta,\theta_0)$ does not depend upon the measure μ in the following sense: if $\bar{\mu}$ is another measure with $\mu \overset{loc}{\ll} \bar{\mu}$ and if $\mathbf{H}(\alpha;\theta,\theta_0)$ and $\mathbf{H}'(\alpha;\theta,\theta_0)$ are the process computed through μ and $\bar{\mu}$, then $\mathbf{H}(\alpha;\theta,\theta_0)$ and $\mathbf{H}'(\alpha;\theta,\theta_0)$ are μ-indistinguishable.

Theorem A.5.2 (Theorem IV.2.1 in Jacod and Shiryaev [23])**.** *Let T be a stopping time. For $\alpha\in[0,1)$ let $\mathbf{H}(\alpha;\theta,\theta_0)$ be any version of the Hellinger process (see IV.1.52 in Jacod and Shiryaev [23] for the case $\alpha=0$). There is equivalence between:*

1. $P_\theta^T \ll P_{\theta_0}^T$;
2. $P_\theta^0 \ll P_{\theta_0}^0$, $P_\theta\left(H(1/2;\theta,\theta_0,T)<\infty\right)=1$ *and* $P_\theta\left(H(0;\theta,\theta_0,T)=0\right)=1$;
3. $P_\theta^0 \ll P_{\theta_0}^0$ *and* $H(\alpha;\theta,\theta_0,T)\overset{P_\theta}{\to} 0$ *as* $\alpha\downarrow 0$.

Remark A.5.3. *We could replace $H(1/2;\theta,\theta_0,T)$ by $H(\beta;\theta,\theta_0,T)$, for any fixed $\beta\in(0,1)$, in the Theorem A.5.2.*

Consider that the space $(\Omega,\mathcal{F})$ is endowed with a càdlàg d-dimensional process $\mathbf{X}=(X^i)_{i\le d}$ and a filtration $\{\mathcal{F}_t; t\ge 0\}$ generated by $\mathbf{X}$.

We denote by $\mu=\mu^X$ the random measure on $\mathbb{R}_+\times\mathbb{R}^d$ associated with the jumps of $\mathbf{X}$. We fix a truncation function h, two triplets (B,C,ν), (B^0,C^0,ν^0) and two initial measures π and π_0 on $(\Omega,\mathcal{F}_0)$. We also consider two probability measures P_θ and P_{θ_0} on $(\Omega,\mathcal{F})$, which are solutions to the martingale problems $s(\mathbf{X}|\pi;B,C,\nu)$ and $s(\mathbf{X}|\pi_0;B^0,C^0,\nu^0)$, respectively.

Although the properties stated in Theorem A.5.1 do characterize the Hellinger process $\mathbf{H}(\alpha;\theta,\theta_0)$, they do not give any "explicit" form to it. In order to obtain a form for the Hellinger Process, we introduce a function $\varphi_\alpha:\mathbb{R}_+^2\to\mathbb{R}_+$ (where for $\alpha\in(0,1)$), defined by

$$\varphi_\alpha(u,v):=\alpha u+(1-\alpha)v-u^\alpha v^{1-\alpha}.$$

We will consider:

1. Let $\mathbf{A}$ be an increasing predictable finite-valued process and let c and c' be two processes taking values in the set of non-negative symmetric $d\times d$-matrices and predictable, such that

$$C=c\cdot A,\quad C'=c'\cdot A,$$

   up tu a $(P_\theta+P_{\theta_0})$-evanescent set;

2. Let λ be a predictable random measure on $\mathbb{R}_+\times\mathbb{R}^d$, such that $(|x|^2+1)*\lambda_t<\infty$ for all $t<\infty$ and that

$$\nu\ll\lambda \text{ and } \nu^0\ll\lambda;$$

3. Let $\mathbf{U} = \{U(t); t \geq 0\}$ be a non-negative predictable function on $\tilde{\Omega}$ such that $\nu = U \cdot \lambda$;

4. We know from (A.2) that $a(\omega, t) := \nu\left(\omega; \{t\} \times \mathbb{R}^d\right) \leq 1$. Up to a $(P_\theta + P_{\theta_0})$-evanescent set, we have
$$a(t) = \int U(t, x)\lambda(\{t\} \times dx) \leq 1;$$

5. Denote by Σ a predictable random set such that
$$\Sigma = \left\{(\omega, t); |h(x)\left(U - U'\right)| * \lambda_t(\omega) < \infty\right\};$$

6. We define the stopping time
$$\tau = \inf\left\{t;\ \text{either } t \notin \Sigma,\ \text{or } C(t) \neq C'(t),\ \text{or } t \in \Sigma \text{ and } \tilde{b} \cdot A(t) + \tilde{B}'(t) \neq 0\right\};$$

Proposition A.5.4 (Corollary IV.3.68 in Jacod and Shiryaev [23])**.** *Suppose that the martingale problem*
$$s\left(\mathbf{X}\middle|\frac{\pi + \pi_0}{2}; \frac{B + B^0}{2}, \frac{C + C^0}{2}, \frac{\nu + \nu^0}{2}\right)$$
has at least one solution and local uniqueness for both problems $s(\mathbf{X}|\pi; B, C, \nu)$ and $s\left(\mathbf{X}|\pi_0; B^0, C^0, \nu^0\right)$. Assume that $\mathbf{H}(1/2)$ does not jump to infinity and that $\tau = \infty$. Then if $\alpha \in (0, 1)$, a version of $\mathbf{H}(\alpha; \theta, \theta_0)$ is
$$H(\alpha; \theta, \theta_0, t) = \frac{\alpha(1-\alpha)}{2}(\beta c \beta)\mathbf{1}_\Sigma \cdot A + \varphi_\alpha(U, U^0) * \lambda + \sum_{s \leq t} \varphi_\alpha\left(1 - a(s), 1 - a^0(s)\right). \tag{A.7}$$

A.6 Exponential Families of Stochastic Process

Consider a family of measures $\{P_\theta; \theta \in \Theta\}$ on a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ with parameter set $\Theta \subset \mathbb{R}^N$ such that the interior int Θ is non-empty. We assume that the filtration $\{\mathcal{F}_t; t \geq 0\}$ satisfies the usual hypothesis.

Definition A.6.1. *The class $\{P_\theta; \theta \in \Theta\}$ is called a* exponential family *on the filtered space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\})$ if there exists a σ-finite measure μ on $(\Omega, \mathcal{F})$ such that, for all $\theta \in \Theta$, $P_\theta \overset{loc}{\ll} \mu$ and*
$$\frac{dP_\theta^t}{d\mu^t} = a(\theta, t)q(t)\exp\left\{\gamma(\theta, t)^\top A(t)\right\},\quad t \geq 0,\quad \theta \in \Theta. \tag{A.8}$$

Considered as a function of θ, this Radon-Nikodym derivative is the *likelihood function* corresponding to the observation of events in $\mathcal{F}_t$.

Definition A.6.2. *A statistical experiment $\{P_\theta; \theta \in \Theta\}$ forms a* curved exponential family *if the likelihood function exists and is of the form*
$$\frac{dP_\theta^t}{dP_0^t} = \exp\left\{\theta^\top A(t) - \kappa(\theta)S(t)\right\} \tag{A.9}$$
where $\kappa : \Theta \to \mathbb{R}_+$ and $A : \Omega \times \mathbb{R}_+ \to \mathbb{R}^N$ is a càdlàg process and $S : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is assumed to be a non-decreasing continuous process with $S(0) = 0$ and $S(t) \to \infty$ P_θ-a.s. as $t \to \infty$ for all $\theta \in \Theta$.

Definition A.6.3. *We say that a function $\kappa : \Theta \to \mathbb{R}_+$ is a* steep function *if for all $\theta_1 \in \Theta \setminus$ int Θ and all $\theta_0 \in$ int Θ,*

$$\frac{d}{ds}\kappa(\theta_s) \to \infty \ \text{ as } s \uparrow 1,$$

where $\theta_s = \theta_0(1-s) + \theta_1 s$, $0 < s < 1$.

Theorem A.6.4 (Theorem 5.2.1 in Küchler and Sϕrensen [33]). *Suppose a curved exponential family of the form (A.9) is given such that $\kappa(\cdot)$ is a steep function and $S(t) > 0$ for $t > 0$ P_θ-a.s., for all $\theta \in \Theta$. Then the maximum likelihood estimator $\hat{\theta}_t$ based on observation in the time interval $[0, t]$ exists and is uniquely given by*

$$\hat{\theta}_t = \dot{\kappa}^{-1}\left(\frac{A(t)}{S(t)}\right)$$

if and only if $\frac{A(t)}{S(t)} \in$ int $\dot{\kappa}($int $\Theta)$.
Suppose $\theta \in$ int Θ. Then under P_θ the maximum likelihood estimator is unique for t sufficiently large and

$$\hat{\theta}_t \overset{t\to\infty}{\longrightarrow} \theta \ \ P_\theta - a.s.$$

Theorem A.6.5 (Theorem 5.2.2 in Küchler and Sϕrensen [33]). *For a curved exponential family of the form (A.9) assume that $\theta \in$ int Θ and there exists an increasing positive non-random function $\phi_\theta(t)$ such that under P_θ*

$$\phi_\theta^{-1}(t)S(t) \to \eta^2(\theta)$$

in probability as $t \to \infty$, where $\eta^2(\theta)$ is a finite random variable such that it is non-negative P_θ-almost sure. Then, under P_θ,

$$\left(S^{-1/2}(t)\left(A(t) - \dot{\kappa}(\theta)S(t)\right), \phi_\theta^{-1}(t)S(t)\right) \to N\left(0, \ddot{\kappa}(\theta)\right) \times F_\theta$$

weakly as $t \to \infty$ conditionally on $[\eta^2(\theta) > 0]$, where F_θ is the conditional distribution of $\eta^2(\theta)$ given $[\eta^2(\theta) > 0]$. Moreover, under P_θ,

$$\left(S^{1/2}\left(\hat{\theta}_t - \theta\right), \phi_\theta^{-1}(t)S(t)\right) \to N\left(0, \ddot{\kappa}^{-1}(\theta)\right) \times F_\theta$$

weakly as $t \to \infty$ conditionally on $[\eta^2(\theta) > 0]$. Moreover,

$$-2\log Q(t) \to \chi^2(k-l) \tag{A.10}$$

weakly as $t \to \infty$ conditionally on $[\eta^2(\theta) > 0]$. Here

$$Q(t) = \frac{sup_{\beta\in B}\mathcal{L}(g(\beta); t)}{sup_{\theta\in\Theta}\mathcal{L}(\theta; t)}$$

is the likelihood ratio test statistic for the hypothesis that the true parameter value θ belongs to an $g(B)$, where $B \subset \mathbb{R}^l$ $(l < k)$ and $g : B \mapsto$ intΘ is a differentiable function for which the matrix $\{\partial g/\partial\beta\}$ has full rank for all $\beta \in B$.

Corollary A.6.6. *Under conditions of Theorem A.6.5,*

$$\phi_\theta^{1/2}(t)\left(\hat{\theta}_t - \theta\right) \to N\left(0, \ddot{\kappa}^{-1}(\theta)\eta^{-2}(\theta)\right)$$

weakly as $t \to \infty$ conditionally on $[\eta^2(\theta) > 0]$.

A more general class of models can be considered, the ones which have likelihood functions

$$\frac{dP_\theta^t}{dP_0^t} = \exp\left\{\sum_{j=1}^{n}\left[\theta_{(j)}^\top A_j(t) - \kappa_j(\theta_{(j)})S_j(t)\right]\right\}, \tag{A.11}$$

where $\theta^\top = \left(\theta_{(1)}^\top, \cdots, \theta_{(n)}^\top\right) \in \Theta_1 \times \cdots \times \Theta_n$, $\Theta_j \in \mathbb{R}^{N_j}$, and int$\Theta_j \neq \emptyset$. For every j we assume that $k_j(\cdot)$ and $S_j(t)$ are one-dimensional and that $A_j(t)$ is a N_j-dimensional càdlàg process. We also assume that for each j, $S_j(t)$ is a non-decreasing continuous process for which $S_j(0) = 0$ and $S_j(t) \to \infty$ P_θ-a.s. as $t \to \infty$.

For this general class of models Küchler and Sϕrensen [33] (Theorem 5.3.1, p.55.) establish the following result.

Theorem A.6.7. *Suppose $S_j(t)$ is strict increasing for the jth component of (A.11). Then the statements of Theorems A.6.4, A.6.5 and A.6.6 hold for this component. If the conditions hold for all $j = 1, \cdots, n$ then (A.10) holds for the hypothesis that $\theta_{(j)}$, $j \in J$, is the true value of these components while other components are unspecified. Here J is an arbitrary subset of $\{1, \cdots, n\}$, $k = \sum_{j \in J} k_j$ in (A.10).*

A.7 Some Basic Concepts in Itô Integration and Martingale Theory

Definition A.7.1. *We say that a martingale is quadratic integrable on $[0, t]$ if its second moment is bounded.*

Theorem A.7.2 (See Theorem 4.7 in Klebaner [27])**.** *Let $\mathbf{X} = \{X(t); t \geq 0\}$ be an adapted process such that $\int_0^t EX^2(s)ds < \infty$. Then*

$$M(t) = \int_0^t X(s)dW(s), \quad 0 \leq t \leq T,$$

is a continuous zero mean square integrable martingale.

Proposition A.7.3 (See Equation (4.25) in Klebaner [27])**.** *Let $M_1(t)$ and $M_2(t)$ be Itô integrals of $X_1(t)$ and $X_2(t)$ regarding the same Wiener process. Then, the quadratic covariation of M_1 and M_2 on $[0, t]$ is given by*

$$[M_1, M_2](t) = \int_0^t X_1(s)X_2(s)ds.$$

Central Limit Theorem

Let us establish some notations before presenting the next theorem. Let $\mathbf{A} = (a_{jk})_{N\times N} \in \mathbb{R}^{N\times N}$ be a positive semi-definite $N \times N$-matrix. Denote $\mathbf{A}^{1/2}$ and det($\mathbf{A}$) its positive semi-definite square root and its determinant, respectively. Let $\mathbf{v}^\top = (v_1, \cdots, v_n) \in \mathbb{R}^N$ be a vector. We denote

$$(\text{diag } \mathbf{A})^\top = (a_{11}, a_{22}, \cdots, a_{NN})$$

the diagonal vector of $\mathbf{A}$ and diag $\mathbf{v}$ the diagonal $N \times N$-matrix with $\mathbf{v}$ as diagonal.

Observe that, with this notation, the identity matrix I_N can be written as $I_N = \text{diag}\,(1, \cdots, 1)^\top$. Furthermore, $\text{diag}(\text{diag}(\mathbf{A}))$ denotes the diagonal matrix that has the same diagonal of $\mathbf{A}$. Remember that the multiples of the identity, i.e.,

$$\{\lambda I_N; \lambda \in \mathbb{R}\}$$

are the center of the group of the quadratic matrix of order $N \in \mathbb{N}$, $N \geq 2$, with non-zero determinant under the product operation. This will allow us to freely commute any matrix with diagonal matrix, when all the elements of the diagonal are the same.

We present below a version of the Central Limit Theorem for N-dimensional martingales in continuous time. This result is the Theorem A.7.7 in Küchler and Sϕrensen [33] and its proof can be found in Küchler and Sϕrensen [34]. It is important to note that the Theorem 2.2 in Crimaldi and Pratelli [8] is a more general version of the Theorem A.7.4 in which they suppress the assumption on the convergence of the third condition.

Theorem A.7.4. *Let* $\mathbf{M} = (\mathbf{M}_1, \cdots, \mathbf{M}_N)^\top = \left\{(M_1(t), \cdots, M_N(t))^\top; t \geq 0\right\}$ *be a N-dimensional square integrable martingale with mean zero and quadratic covariation matrix $[M]$. Let $H(t)$ denote the covariance matrix of $M(t)$, i.e.,*

$$H(t) = E\left[M(t)M(t)^\top\right].$$

Assume that there exists a family of invertible non-random $N \times N$-matrices $\{K(t); t \geq 0\}$ such that as $t \to \infty$ we have $K(t) \to 0$ and

1. $\sum_{j=1}^{N} |K_{jk}(t)|\, E\left[\sup_{s \leq t} |\Delta M_k(t)|\right] \to 0$, $k = 1, \cdots, N$;
2. $K(t)[M](t)K(t)^\top \to W$ *in probability,*
   where W is a random positive semi-definite matrix satisfying $\mathbb{P}\,(\det W > 0) > 0$;
3. $K(t)H(t)K(t)^\top \to \Sigma$,
   where Σ is a positive definite (deterministic) matrix.

   Then

   $$\left(K(t)M(t), K(t)[M](t)K(t)^\top\right) \to \left(W^{1/2}Z, W\right)$$

   and, conditionally on $[\det W > 0]$,

   $$W^{-1/2}K(t)M(t) \to Z$$

   in distribution as $t \to \infty$, where Z is an N-dimensional standard normal distributed random vector independent of W.

Law of Larger Numbers

Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ be a filtered probability space.

Theorem A.7.5 (Liptser [38])**.** *Let $\mathbf{M} = \{M(t); t \geq 0\}$ be a locally square integrable martingale with $M(0) = 0$ and suppose that $\mathbf{A} = \{A(t); t \geq 0\}$ is a predictable, non-decreasing and right-continuous process with $A(0) = 0$. Define*

$$B(t) = \int_0^t (1 + A(s))^{-2}\, d[M](s).$$

Then, as $t \to \infty$,

$$\frac{M(t)}{A(t)} \xrightarrow{a.s.} 0,$$

on $[A(\infty) = \infty] \cap [B(\infty) < \infty]$.

Uniform Law of Large Numbers

Let $\Theta \subset \mathbb{R}^N$ be a compact set. Consider a family of real processes $\{M(\theta, t); \theta \in \Theta \text{ and } t \geq 0\}$ on a probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t; t \geq 0\}, \mathbb{P})$ such that for all θ the process $\{M(\theta, t); t \geq 0\}$ is a continuous local martingale starting at zero. Denote by $[M(\theta)](t)$ the quadratic variation of $M(\theta, t)$ and by $[M(\theta, \theta')](t)$ that of $M(\theta, t) - M(\theta', t)$.

Theorem A.7.6 (Theorem 2 in Loukianova and Loukianov [40]). *Suppose that there exists a constant* $\delta \in (0, 1]$ *and a continuous increasing process* $V(t) > 0$ *such that* $V(\infty) = \infty$ *a.s. and for all* $(\theta, \theta', t) \in \Theta^2 \times \mathbb{R}_+$

$$[M(\theta, \theta')](t) \leq V(t)|\theta - \theta'|^{2\delta} \ \ a.s.$$

Suppose also that there exist $\theta' \in \Theta$ *such that*

$$\limsup_{t\to\infty} \frac{[M(\theta')](t)}{V(t)} < \infty \ \ a.s.$$

Then there exists a continuous in θ *modification* $\tilde{M}(\theta, t)$ *of* $M(\theta, t)$ *such that*

$$\lim_{t\to\infty} \sup_{\theta\in K} \frac{\left|\tilde{M}(\theta, t)\right|}{[M(\theta)](t)} = 0 \ \ a.s.$$

for any compact $K \subset \Theta$ *satisfying that for all* $\theta \in K$

$$\liminf_{t\to\infty} \frac{[M(\theta)](t)}{V(t)} > 0 \ \ a.s.$$

Burkholder-Davis-Gundy Inequalities

Proposition A.7.7 (Burkholder-Davis-Gundy Inequalities). *For every* $0 < p < \infty$, *there exist two constants* c_p *and* C_p *such that, for all local martingale* $\mathbf{M} = \{M(t);\ t \geq 0\}$ *vanishing at zero, for any stopping time* T *and any bounded predictable process* $\mathbf{H} = \{H(t);\ t \geq 0\}$

$$\begin{aligned} c_p E\left[\left(\int_0^T H^2(s)d[M](s)\right)^{p/2}\right] &\leq E\left[\sup_{t\leq T}\left|\int_0^t H(s)dM(s)\right|^p\right] \\ &\leq C_p E\left[\left(\int_0^T H^2(s)d[M](s)\right)^{p/2}\right]. \end{aligned} \tag{A.12}$$

Proof. See Revuz and Yor [57, Chapter IV, Section 4, pp.160-170].

A.8 M-Estimators

The next result gives us a sufficient condition for the consistency in $P_{\theta_0}-$probability of $M-$estimators.

Consider $\Theta \subset \mathbb{R}^N$ such that (Θ, d) is a metric space. Given an arbitrary random function $\theta \mapsto M_n(\theta)$, consider estimators $\hat{\theta}_n$ that nearly maximize M_n, that is

$$M_n(\hat{\theta}_n) \geq \sup_{\theta\in\Theta} M_n(\theta) - o_{P_{\theta_0}}(1)\cdot$$

Theorem A.8.1 (Theorem 5.7 in Vaart [65]). *Let M_n be random functions and let M be a fixed function of θ such that for every $\varepsilon > 0$*

1. $\sup_{\theta \in \Theta} |M_n(\theta) - M(\theta)| \to 0$ *in P_{θ_0}–probability;*

2. $\sup_{\theta; d(\theta,\theta_0) \geq \varepsilon} M_n(\theta) < M(\theta_0)$.

Then any sequence of estimators $\hat{\theta}_n$ with $M_n(\hat{\theta}_n) \geq M_n(\theta_0) - o_{P_{\theta_0}}(1)$ converges in P_{θ_0}–probability to θ_0.

Bibliography

[1] W. Alcântara. *Uma Nova Classe de Soluções da Equação de Langevin Generalizada e sua Aplicação na Modelagem do Cupom Cambial*. PhD thesis, Universidade de Brasília, 2019.

[2] D. Applebaum. *Levy processes and stochastic calculus*. Cambridge studies in advanced mathematics 93. Cambridge University Press, 1 edition, 2004.

[3] O. E. Barndorff-Nielsen and N. Shephard. Non-gaussian ornstein-uhlenbeck-based models and some of their uses in financial economics. *Journal of the Royal Statistical Society, Series B*, 63:167–241, 2001.

[4] J. M. Benke. *Asymptotic inference for linear stochastic differential equations with time delay*. PhD thesis, szte, 2018.

[5] J. M. Benke and G. Pap. Asymptotic inference for a stochastic differential equation with uniformly distributed time delay. *Journal of Statistical Planning and Inference*, 167:182 – 192, 2015.

[6] J. M. Benke and G. Pap. One-parameter statistical model for linear stochastic differential equation with time delay. *Statistics*, 51(3):510–531, 2017.

[7] P. Billingsley. *Convergence of probability measures*. Wiley series in probability and statistics. Probability and statistics section. Wiley, 2nd ed edition, 1999.

[8] I. Crimaldi and L. Pratelli. Convergence results for multivariate martingales. *Stochastic Processes and their Applications*, 115(4):571 – 577, 2005.

[9] I. DiTerlizzi, F. Ritort, and M. Baiesi. Explicit solution of the generalised langevin equation. *Journal of Statistical Physics*, 181(5):1609–1635, 2020.

[10] C. C. Dorea. Stopping rules for a random optimization. *SIAM Journal on Control and Optimization*, 28:841–850, 1990.

[11] W. Feller. *An introduction to probability theory and its applications*, volume Volume 1. Wiley, 3 edition, 1968.

[12] R. F. Fox. The generalized langevin equation with gaussian fluctuations. *Journal of Mathematical Physics*, 18(12):2331–2335, 1977.

[13] A. Gloter, D. Loukianova, and H. Mai. Jump filtering and efficient drift estimation for lévy-driven sdes. *Ann. Statist.*, 46(4):1445–1480, 2018.

[14] A. A. Gushchin and U. Küchler. On parametric statistical models for stationary solutions of affine stochastic delay differential equations. *Mathematical Methods of Statistics*, 12(1):31—-61, 2003.

[15] A. A. Gushchin and U. Küchler. On estimation of delay location. *Statistical inference for stochastic processes*, 14(3):273–305, 2011.

[16] A. A. Gushchin and U. Küchler. Asymptotic inference for a linear stochastic differential equation with time delay. *Bernoulli*, 5(6):1059–1098, 1999.

[17] A. A. Gushchin and U. Küchler. On stationary solutions of delay differential equations driven by a lévy process. *Stochastic Processes and their Applications*, 88(2):195 – 211, 2000.

[18] J. Hájek. A characterization of limiting distributions of regular estimates. *eitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14:323 – 330, 1970.

[19] J. Hájek. Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Theory of Statistics*, pages 175–194, Berkeley, Calif., 1972. University of California Press.

[20] Y. Hu and H. Long. Least squares estimator for ornstein–uhlenbeck processes driven by α-stable motions. *Stochastic Processes and their Applications*, 119(8):2465 – 2480, 2009.

[21] J. E. Hutton and P. I. Nelson. Interchanging the order of differentiation and stochastic integration. *Stochastic Processes and their Applications*, 18(2):371 – 377, 1984.

[22] S. K. I.A. Ibragimov, R.Z. Has'minskii. *Statistical estimation. Asymptotic theory*. Springer, 1st edition, 1981.

[23] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*. Grundlehren der mathematischen Wissenschaften 288. Springer Berlin Heidelberg, 2nd ed edition, 1987.

[24] D. Kannan. On the generalized langevin equation. *Journal of Mathematical and Physical Sciences*, 11(1):1–24, 1977.

[25] D. Kannman and A. Bharucha-Reid. Random integral equation formulation of a generalized langevin equation. *Journal of Statistical Physics*, 5(3):209–233, 1972.

[26] S. Ken-Iti. *Lévy processes and infinitely divisible distributions*. Cambridge university press, 1999.

[27] F. C. Klebaner. *Introduction to stochastic calculus with applications*. Imperial College Press, 2 edition, 2005.

[28] A. Kohatsu-Higa, E. Nualart, and N. K. Tran. Lan property for an ergodic diffusion with jumps. *Statistics*, 51(2):419–454, 2017.

[29] R. Kubo. The fluctuation-dissipation theorem. *Reports on Progress in Physics*, 29(1):255–284, jan 1966.

[30] U. Küchler, M. Sørensen, et al. Statistical inference for discrete-time samples from affine stochastic delay differential equations. *Bernoulli*, 19(2):409–425, 2013.

[31] A. M. Kulik. Exponential ergodicity of the solutions to sde's with a jump noise. *Stochastic Processes and their Applications*, 119(2):602 – 632, 2009.

[32] Y. A. Kutoyants. *Statistical inference for ergodic diffusion processes*. Springer Science & Business Media, 2004.

[33] U. Küchler and M. Sorensen. *Exponential Families of Stochastic Processes*. Springer Series in Statistics. Springer, 1 edition, 1997.

[34] U. Küchler and M. Sørensen. A note on limit theorems for multivariate martingales. *Bernoulli*, 5(3):483–493, 06 1999.

[35] L. LeCam. Locally asymptotically normal families of distributions. *University of California Publications in Statistics*, 3:37 – 98, 1960.

[36] L. LeCam and G. L. Yang. *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer-Verlag New York, 2 edition, 2000.

[37] S. H. Lim, J. Wehr, and M. Lewenstein. Homogenization for generalized langevin equations with applications to anomalous diffusion. In *Annales Henri Poincaré*, pages 1–59. Springer, 2020.

[38] R. Liptser. A strong law of large numbers for local martingales. *Stochastics*, 3(1-4):217–228, 1980.

[39] Y. Liu, E. Nualart, and S. Tindel. Lan property for stochastic differential equations with additive fractional noise and continuous time observation. *Stochastic Processes and their Applications*, 129(8):2880 – 2902, 2019.

[40] D. Loukianova and O. Loukianov. Uniform law of large numbers and consistency of estimators for harris diffusions. *Statistics & Probability Letters*, 74(4):347 – 355, 2005.

[41] M. Magdziarz and A. Weron. Ergodic properties of anomalous diffusion processes. *Annals of Physics*, 326(9):2431–2443, 2011.

[42] H. Mai. *Drift estimation for jump diffusions*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2012.

[43] H. Mai. Efficient maximum likelihood estimation for lévy-driven ornstein–uhlenbeck processes. *Bernoulli*, 20(2):919–957, 05 2014.

[44] C. Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296, 2009.

[45] C. Mancini. The speed of convergence of the threshold estimator of integrated variance. *Stochastic Processes and their Applications*, 121(4):845 – 855, 2011.

[46] H. Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes and their Applications*, 117(1):35 – 56, 2007.

[47] H. Masuda. Erratum to: “ergodicity and exponential β -mixing bound for multidimensional diffusions with jumps” [stochastic process. appl. 117 (2007) 35–56]. *Stochastic Processes and their Applications*, 119(2):676 – 678, 2009.

[48] S. A. McKinley and H. D. Nguyen. Anomalous diffusion and the generalized langevin equation. *SIAM Journal on Mathematical Analysis*, 50(5):5119–5160, 2018.

[49] A. V. Medino, S. R. Lopes, R. Morgado, and C. C. Dorea. Generalized langevin equation driven by lévy processes: A probabilistic, numerical and time series based approach. *Physica A: Statistical Mechanics and its Applications*, 391(3):572 – 581, 2012.

[50] R. Morgado, F. A. Oliveira, G. G. Batrouni, and A. Hansen. Relation between anomalous and normal diffusion in systems with memory. *Physical Review Letters*, 89(10):100601, 2002.

[51] H. Mori. Transport, Collective Motion, and Brownian Motion*). *Progress of Theoretical Physics*, 33(3):423–455, 03 1965.

[52] M. Ottobre and G. A. Pavliotis. Asymptotic analysis for the generalized langevin equation. *Nonlinearity*, 24(5):1629, 2011.

[53] G. A. Pavliotis, G. Stoltz, and U. Vaes. Scaling limits for the generalized langevin equation. *Journal of Nonlinear Science*, 31(1):1–58, 2021.

[54] P. E. Protter. *Stochastic integration and differential equations*. Applications of mathematics 21 0172-4568. Springer, 2nd ed edition, 2004.

[55] M. Reiß. *Nonparametric estimation for stochastic delay differential equations*. PhD thesis, Humboldt-Universität zu Berlin, Mathematisch-Naturwissenschaftliche Fakultät II, 2002.

[56] M. Reiß. Adaptive estimation for affine stochastic delay differential equations. *Bernoulli*, 11(1):67–102, 2005.

[57] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*. Grundlehren der mathematischen Wissenschaften. Springer, 3rd edition, 1998.

[58] G. Samorodnitsky and M. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance (Stochastic Modeling Series)*. Chapman and Hall/CRC, 1 edition, 1994.

[59] F. F. T. Santos. *Classes de Soluções para a Equação de Langevin Generalizada*. PhD thesis, Departamento de Matemática, Universidade de Brasília (UnB), 2011.

[60] J. Slezak. *Langevin equation and fractional dynamics*. PhD thesis, Wroclaw University of Science and Technology, 2018.

[61] J. Stein. *Estimação e Previsão em Processos Advindos da Solução da Equação de Langevin Generalizada com Ruído α-Estável*. PhD thesis, Universidade Federal do Rio Grande do Sul, 2015.

[62] J. Stein, S. R. C. Lopes, and A. V. Medino. Continuous processes derived from the solution of generalized langevin equation: theoretical properties and estimation. *Journal of Statistical Computation and Simulation*, 86(14):2819–2845, 2016.

[63] N. K. Tran. *LAN property for jump-diffusion processes with discrete observations via Malliavin calculus*. PhD thesis, Université Paris 13, 2014.

[64] N. K. Tran. Lan property for an ergodic ornstein–uhlenbeck process with poisson jumps. *Communications in Statistics - Theory and Methods*, 46:7942 – 7968, 2015.

[65] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1 edition, 1998.

[66] Y. Zhu and D. Venturi. Generalized langevin equations for systems with local interactions. *Journal of Statistical Physics*, pages 1–31, 2020.

[67] R. Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, 2001.