Drift estimation for jump diffusions: time-continuous and high-frequency observations

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. Rer. Nat.
im Fach Mathematik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät II
Humboldt-Universität zu Berlin

von

Dipl.-Math. Hilmar Mai

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät II:
Prof. Dr. Elmar Kulke

Gutachter:
1. Prof. Dr. Uwe Küchler (Betreuer)
2. Prof. Dr. Markus Reiß (Betreuer)
3. Prof. Dr. Vladimir Spokoiny
4. Prof. Dr. Michael Sørensen

eingereicht am: 24.04.2012
Tag der mündlichen Prüfung: 28.09.2012

Abstract

The problem of parametric drift estimation for a Lévy-driven jump diffusion process is considered in two different settings: time-continuous and high-frequency observations. The goal is to develop explicit maximum likelihood estimators for both observation schemes that are efficient in the Hájek-Le Cam sense.

In order to develop a maximum likelihood approach the absolute continuity and singularity problem for the induced measures on the path space is discussed. For varying drift parameter we obtain locally equivalent measures when the driving Lévy process has a Gaussian component. The likelihood function based on time-continuous observations can be derived explicitly and leads to explicit maximum likelihood estimators for several popular model classes. We consider Ornstein-Uhlenbeck type, square-root and linear stochastic delay differential equations driven by Lévy processes in detail and prove strong consistency, asymptotic normality and efficiency of the likelihood estimators in these models.

The appearance of the continuous martingale part of the observed process under the dominating measure in the likelihood function leads to a jump filtering problem in this context, since the continuous part is usually not directly observable and can only be approximated and the high-frequency limit. This leads to the question how the jumps of the driving Lévy process influence the estimation error. We show that when the continuous part can only be recovered up to some small jumps the estimation error is proportional to the jump intensity of these small jumps. Hence, efficient jump filtering becomes an important task before inference on the drift can be undertaken. As a side result we obtain that least squares estimation is inefficient when jumps are present.

In the second part of this thesis the problem of drift estimation for discretely observed processes is considered. The estimators are constructed from discretizations of the time-continuous maximum likelihood estimators from the first part, where the continuous martingale part is approximated via a threshold technique. Here the jump activity of the Lévy process plays a crucial role for the asymptotic analysis of the estimators. We consider first the case of finite activity and show that under suitable conditions on the behavior of small jumps and the observation frequency the drift estimator attains the efficient asymptotic distribution that we have derived in the first part. Based on these results we prove asymptotics normality and efficiency for the drift estimator in the Ornstein Uhlenbeck type model also for infinite jump activity. In the course of the proof we show that the continuous part of a jump diffusion can be recovered in the high frequency limit even when the observation horizon growth to infinity and the process has infinitely many small jumps in every finite time interval.

Finally, the finite sample behavior of the estimators is investigated on simulated data. When the assumption of high-frequency observations is reasonable the theoretical results are confirmed. We find also that the maximum likelihood approach clearly outperforms the least squares estimator when jumps are present and that the efficiency gap between both techniques becomes even more severe with growing jump intensity.

Zusammenfassung

Das Ziel dieser Arbeit ist die Entwicklung eines effizienten parametrischen Schätzverfahrens für den Drift einer durch einen Lévy-Prozess getriebenen Sprungdiffusion. Zunächst werden zeit-stetige Beobachtungen angenommen und auf dieser Basis eine Likelihoodtheorie entwickelt. Dieser Schritt umfasst die Frage nach lokaler Äquivalenz der zu verschiedenen Parametern auf dem Pfadraum induzierten Maße. Es zeigt sich, dass lokale Äquivalenz vorliegt sobald der treibende Lévy-Prozess einen Gauß'schen Anteil besitzt. In diesem Fall kann die zugehörige Likelihood-Funktion explizit angegeben werden, so dass für einige in Anwendungen populären Modelle ein expliziter Maximum-Likelihood-Schätzer entwickelt werden kann. Wir diskutieren in dieser Arbeit Schätzer für Prozesse vom Ornstein-Uhlenbeck-Typ, Cox-Ingersoll-Ross Prozesse und Lösungen linearer stochastischer Differentialgleichungen mit Gedächtnis im Detail und zeigen starke Konsistenz, asymptotische Normalität und Effizienz im Sinne von Hájek und Le Cam für den Likelihood-Schätzer.

In Sprungdiffusionsmodellen ist die Likelihood-Funktion eine Funktion des stetigen Martingalanteils des beobachteten Prozesses, der im Allgemeinen nicht direkt beobachtet werden kann. Wenn nun nur Beobachtungen an endlich vielen Zeitpunkten gegeben sind, so lässt sich der stetige Anteil der Sprungdiffusion nur approximativ bestimmen. Diese Approximation des stetigen Anteils ist ein zentrales Thema dieser Arbeit und es wird uns auf das Filtern von Sprüngen führen. Um den Einfluss der Sprünge auf den Schätzfehler besser zu verstehen, nehmen wir nun an, dass nur große Sprünge entfernt werden können. Unter diesen durch Sprünge gestörten Daten zeigt sich, dass der Schätzfehler des Maximum-Likelihood-Schätzers proportional zur Sprungintensität ist, so dass die Entfernung des Sprunganteils aus den Daten wichtig für die Effizienz des Schätzers wird. Als Korollar dieser Untersuchungen erhalten wir, dass der Kleinste-Quadrate-Schätzer in Modellen mit Sprüngen ineffizient ist.

Der zweite Teil dieser Arbeit untersucht die Schätzung des Drifts, wenn nur diskrete Beobachtungen gegeben sind. Dabei benutzen wir die Likelihood-Schätzer aus dem ersten Teil und approximieren den stetigen Martingalanteil durch einen sogenannten Sprungfilter. Hierfür spielt die Sprungaktivität des treibenden Lévy-Prozesses eine entscheidende Rolle. Wir untersuchen zuerst den Fall endlicher Aktivität und zeigen, dass die Driftschätzer im Hochfrequenzlimes die effiziente asymptotische Verteilung erreichen. Darauf aufbauend beweisen wir dann im Falle unendlicher Sprungaktivität asymptotische Effizienz für den Driftschätzer im Ornstein-Uhlenbeck Modell. Der Beweis beinhaltet als wesentlichen Schritt, dass der stetige Anteil einer Sprungdiffusion aus Hochfrequenzdaten rekonstruiert werden kann, selbst wenn der Beobachtungshorizont gegen Unendlich geht und der treibende Prozess unendlich viele kleine Sprünge pro Zeitintervall aufweist.

Im letzten Teil werden die theoretischen Ergebnisse für die Schätzer auf endlichen Stichproben aus simulierten Daten geprüft. Es zeigt sich, dass ab einer gewissen Beobachtungsdichte der stochastische Fehler den Diskretisierungsfehler dominiert und in diesem Bereich die Annahme hochfrequenter Daten sinnvoll erscheint. Daneben wird auch die Ineffizienz des Kleinste-Quadrate-Schätzers im Vergleich der Standardabweichungen beider Schätzverfahren deutlich und es fällt auf, dass der Effizienzgewinn des Maximum-Likelihood-Schätzers durch den Sprungfilter mit steigender Sprungintensität weiter zunimmt.

Contents

My deepest gratitude and appreciation are due to my supervisors Uwe Küchler and Markus Reiß for their enduring support and mathematical guidance in the course of this work. I would like to thank all my colleagues, in particular Markus, Johanna, Jakob and Mathias, for all the fruitful mathematical discussions and for making our group in Berlin what it was. The patience and encouragement of my parents and my brother have helped me in many difficult situations and I feel deeply indebted to them.

1 Introduction

The study of jump diffusion processes in probability theory was initiated already in the early work of Kolmogoroff [1931] and Feller [1940] on Markov processes. Later Itô's theory of stochastic integration opened up another perspective on jump diffusions as solutions of stochastic differential equations. Since then this class of processes has been employed to describe complex dynamics in all kinds of applications (we will give some examples later on). In particular during the last two decades they have become an inevitable tool in stochastic modeling. But what are the main reasons for the recent interest in this family of continuous time processes from an applied perspective?

Besides their analytical tractability and flexibility to describe many kinds of complex dynamics the main reason might be that we have seen a dramatic change in the availability of data recently. In classical statistics a dynamic process in time is usually represented by a time series, i.e. a stochastic process in discrete time. This modeling framework was for a long time reasonable, since the available data was inherently discrete in the sense that measurements were not frequent enough to employ time-continuous models such as jump diffusions. This situation changed rapidly over the last two decades with the emergence of computer-aided measurements in e.g. physics, biology and physiology, but also electronic trading in finance such that nowadays large sets of so-called high-frequency data are available that make statistical modeling via stochastic processes in continuous time feasible.

The term jump diffusion already reveals that such a process X may be decomposed as the sum of a continuous diffusion component M and a component J that evolves purely by jumps. Here the diffusive part M will be of the form

$$M_t = M_0 + \int_0^t \delta(s, X_s)\, ds + \int_0^t \gamma(s, X_s)\, dW_s, \quad t \in \mathbb{R}_+,$$

where W is a Wiener process. Our aim in this work is to infer the drift or trend function δ from observations of X. In this model the function δ is measured under two quite different nuisance terms: W and J. Each of these noise components on its own already leads to a mathematically challenging estimation problem and for us it will be crucial to understand the interplay of both noise terms in order to develop an estimation approach that recovers δ efficiently.

Statistical model and estimation problem

The main goal of this thesis is to provide a parametric estimation approach for the drift of a jump diffusion process with Lévy noise. A jump diffusion means here the strong solution X to the stochastic integral equation

$$X_t = X_0 + \int_0^t \delta(\theta, s, X_s)\, ds + \int_0^t \gamma(s, X_{s-})\, dL_s, \quad t \in \mathbb{R}_+. \tag{1.1}$$

with initial value $X_0 \in \mathbb{R}$. We suppose that X is defined on a filtered probability space denote by $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$. The drift coefficient δ is parametrized by an unknown $\theta \in \Theta$ and we impose Lipschitz and linear growth conditions on δ and γ such that (1.1) exhibits a unique strong solution. The driving process L is assumed to be a Lévy process with Lévy-Khintchine characteristics (b, σ^2, μ). This triplet of characteristics determines the law of L uniquely via the Lévy-Khintchine formula, which implies a decomposition, the so-called Lévy-Itô decomposition, of L into a linear drift with slope b, a Wiener process W with variance $E[W_t] = \sigma^2 t$ and a jump component that is fully described by the so-called Lévy measure μ.

The jump diffusion process (1.1) incorporates many widely used models from applications. There is such a vast amount of literature such that we can only mention some examples here. The first application in finance was developed by Merton [1976] in the context of option pricing. In the literature on stochastic volatility two well known examples are Bates [1996] and Barndorff-Nielsen and Shephard [2001] (cf. also Cont and Tankov [2004a]). A more general discussion of affine jump-diffusions in finance with a focus on spectral methods for option pricing and estimation was given in Duffie et al. [2000]. In neuroscience the neuronal membrane potential has been represented by a jump diffusion in Lansky and Lanska [1987] and Jahn et al. [2011], where the jump component describes the spiking behavior of the neurons.

For the statistical analysis of jump diffusion models several authors have investigated calibration from option pricing data in financial applications. These references include for example Cont and Tankov [2004b] and Belomestny and Schoenmakers [2011], where spectral estimation techniques based on the empirical Fourier transform were used. Bandi and Nguyen [2003] considered non-parametric estimation of the conditional infinitesimal moments and proved consistency of kernel estimates in this setting. There is also a growing literature on statistical inference in the more general setting of Itô semimartingales as for example in Clement et al. [2011] and the references therein.

When time-continuous observations $(X_t)_{t\in[0,T]}$ are given a natural question is, which characteristics of (1.1) can be identified. It is well known that when $T > 0$ is fixed the quadratic variation of X is known and hence also the integrated volatility $\int_0^T \gamma(s, X_s)^2\, ds$. The situation is completely different for the drift, which cannot be identified in general when T is fixed. However, when $T \to \infty$ the drift is identifiable in the limit. Therefore, we will consider here an observation scheme with with growing time horizon. When X is discretely observed on time points $0 = t_1 < \ldots < t_n = T$ there is a well developed theory for estimation of the volatility under high-frequency observations, i.e. $\Delta_n = \max_{1\leq i\leq n-1}\{t_{i+1} - t_i\} \downarrow 0$, even if X is corrupted by an additive noise (see e.g. Bibinger [2011]). For identification of the drift an observation scheme is needed that satisfies $T_n = T \to \infty$ and also $\Delta_n \downarrow 0$ as $n\ to \infty$.

Let us next sketch our estimation approach. Since every Lévy process has a modification with càdlàg paths, we can assume here that the paths of X lie in the Skorokhod space $D[0, \infty)$ of càdlàg functions on $[0, \infty)$. Under certain conditions the measures P^θ for $\theta \in \Theta$ induced by X on $D[0, \infty)$ are locally equivalent and the Radon-Nikodym derivative or likelihood function is given by

$$\frac{dP_t^\theta}{dP_t^0} = \exp\left[\int_0^t \gamma(s, X_{s-})^{-2}\delta(\theta, s, X_{s-})\, dX_s^c - \frac{1}{2}\int_0^t \beta(\theta, s, X_s)^2\gamma(s, X_{s-})^{-1}\, ds\right],$$

where P_t^θ is the restriction of P^θ to $\mathcal{F}_t$ and X^c denotes the continuous martingale part of X under P^0. There is a well developed theory for parameter estimation for diffusions without jumps driven by a Wiener process. A comprehensive overview for the ergodic case is provided in Kutoyants [2004]. A likelihood theory for jump diffusions under time-continuous observations can be found in Sørensen [1991]. In this thesis we will expand these results and develop an estimation approach for discretely observed X. For jump processes this step involves new mathematical challenges, since X^c is in general unknown and has to be recovered. We will call this approximation of X^c a jump filtering problem.

The appearance of the continuous component in the likelihood function leads to a central problem for statistical inference from jump process models: the separation of the continuous and the jump part. If the statistician is interested in properties of the continuous component, then the jumps can be seen as a noise that has to be filtered or smoothed out before inference on the continuous part can be undertaken. This type of problem occurs in the present work or in the context of volatility estimation in Mancini [2009] or Cont and Mancini [2011], where the integrated volatility of the continuous part is estimated by means of realized quadratic variations and thus the quadratic variation that stems from the jump component has to be removed. The second type of problem occurs when characteristics of the jump behavior are of interest. In Aït-Sahalia and Jacod [2012] for example generalized Blumenthal-Getoor indices for Itô semimartingales are estimated by approximating the number of jumps larger than a certain threshold, which is then used to estimate the activity of small jumps when the threshold tends to zero.

As a main example for such a jump filtering problem we shall consider an Ornstein-Uhlenbeck type process X defined via

$$dX_t = -aX_t\, dt + dL_t, \quad t \in \mathbb{R}_+, \quad X_0 = x \in \mathbb{R},$$

for $a \in \mathbb{R}$ unknown. The recent interest in this class of processes has been mainly stimulated by Barndorff-Nielsen and Shephard [2001] in the context of stochastic volatility modeling in finance. Nonparametric estimation of the Lévy measure of L was considered in Jongbloed et al. [2005]. When L is a subordinator Brockwell et al. [2007] applied time series techniques to infer the drift parameter based on equidistant observations. For purely α-stable L Hu and Long [2009] proposed least squares estimation and proved convergence to a stable limiting distribution.

For time-continuous observations $(X_t)_{t\in[0,T]}$ the efficient maximum likelihood estimator for a is explicitly given by

$$\hat{a}_T = -\frac{\int_0^T X_s dX_s^c}{\int_0^T X_s^2 ds}. \tag{1.2}$$

In applications time-continuous observations are usually not available such that the question arises, how the continuous martingale part can be recovered from discrete observations

$$X_{t_1}, \dots, X_{t_n} \text{ for } 0 = t_1 < \dots < t_n = T_n.$$

Can we identify increments of X that contain jumps? When high-frequency data is available it turns out that this is indeed possible under restriction on the intensity of small jumps by deleting increments of the process $\Delta_i X = X_{t_{i+1}} - X_{t_i}$ that are large relative to the threshold $(t_{i+1} - t_i)^\beta$ for suitably chosen threshold exponent $\beta \in (0, 1/2)$. This leads to the following estimator with jump filter

$$\bar{a}_n = -\frac{\sum_{i=0}^{n-1} X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \le \Delta_n^\beta\}}}{\sum_{i=0}^{n-1} X_{t_i}(t_{i+1} - t_i)}, \tag{1.3}$$

where $\Delta_n = \max_{1 \le i \le n-1}\{t_{i+1} - t_i\}$. In the context of volatility estimation threshold techniques were first employed by Mancini [2009]. Also the recent book by Jacod and Protter [2012] provides detailed discussions of such separation problems between continuous and jump component for Itô-semimartingales based on high-frequency observations. In contrast to our discussion these authors have consider the case of a fixed observation horizon $T_n = T < \infty$ for all $n \in \mathbb{N}$.

One of the main problems considered in this thesis is the question under what conditions on β, the observation scheme and the Lévy measure μ, does $\sum_{i=0}^{n-1} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \le \Delta_n^\beta\}}$ approximate the continuous martingale part X^c such that the drift estimator $\bar{a}_n$ attains the same asymptotic distribution as $\hat{a}_T$ and is efficient? We will tackle this question in two steps. First we assume that L has only jumps of finite activity such that in principle it is possible to identify all jumps when the observation frequency is high enough. In the second most challenging step we generalize these results to the case of possibly infinite jump activity, where we find that if the jump component has an α-stable like behavior for the small jumps we can indeed choose the threshold exponent β such that the continuous part can be identified in the limit and $\bar{a}_n$ attains the efficient asymptotic distribution. This result requires a fine estimate for the Markov generator of the jump component of L and a sophisticated analysis of the convergence behavior of each component of X under thresholding.

Main results and a guideline for the reader

This thesis can be divided into two main parts. In the first part we develop maximum likelihood estimators for the drift based on time-continuous observations for several jump diffusion processes that lead to an explicit estimator and prove asymptotic properties such as consistency, asymptotic normality and efficiency. The second part is devoted to the problem of estimating the drift from discrete observations.

The first part in Chapter 3 and 4 lies the groundwork for the estimation theory from discrete observations in Chapter 5 and 6. We will build on these results in two ways. First of all the discrete estimators will be constructed from their continuous analogs via discretization and jump filtering. Secondly, the asymptotics of the continuous case will serve as a benchmark for the discrete case in the sense that discrete observations cannot be more informative than the fully observed process. Thus the efficiency bounds for the asymptotic variances from Chapter 4 hold also for estimators based on discrete data. Another way of comparing both observation schemes is to look at their limits. Since the high frequency scheme converges as $\Delta_n \to 0$ to the time-continuous scheme, it follows that in the limit both experiments are equally informative such that efficiency bounds carry over from one to the other.

In Chapter 3 we discuss the absolute continuity problem for the measures $(P^\theta)_{\theta\in\Theta}$ induced by the jump diffusion X on the path space $D[0,\infty)$ for different parameters. When absolute continuity holds the likelihood function is known explicitly. These results are based on Sørensen [1991]. From the general results for solutions of (1.1) we specialize then on three specific models that play a major role in applications. The first model will be the class of Ornstein-Uhlenbeck type processes, for which we give an independent and worked out proof of the absolute continuity of solution measures when the driving Lévy process L has a Gaussian component. Our goal is to exemplify how the general theory of absolute continuity and singularity problems for semimartingales, that was developed in Jacod and Memin [1976], applies in our setting. The main tool here is the Hellinger process corresponding to the family $(P^\theta)_{\theta\in\Theta}$. We also discuss the role of the continuous martingale part and its behavior under changes of measure, since this will be crucial later for the investigation of the maximum likelihood estimator. The second example is the class of Lévy-driven square root processes. This class enjoys great popularity in mathematical finance, owing to the fact that it stays non-negative under certain conditions on the driving process. After the work of Cox et al. [1985] they also became known as Cox-Ingersoll-Ross processes. In the last model we exemplify that the likelihood approach also works in the non-Markovian setting of stochastic delay differential equations. Here we use results from Küchler and Sørensen [1989] to derive the likelihood function for solutions of stochastic delay equations driven by Lévy processes.

In Chapter 4 we start by defining a maximum likelihood estimator for the general jump diffusion model (1.1). In this generality the likelihood equation has no explicit solution such that numerical methods have to be applied. In Section 4.2 we develop a detailed asymptotic theory for the maximum likelihood estimator (1.2) for Ornstein-Uhlenbeck type processes X. We prove strong consistency, asymptotic normality and that the statistical experiment is locally asymptotically normal, i.e. that it behaves locally like a Gaussian shift experiment. This property then implies asymptotic efficiency in the sense of the Hájek-Le Cam convolution theorem. Then we investigate the influence of the jumps of the driving Lévy process on the estimation error. Theorem 4.2.10 states that when the continuous martingale part in (1.2) is replaced by $X^c + X^j$, where X^j is a pure jump Lévy process with Lévy measure $\mathbf{1}_{[-\epsilon,\epsilon]}\mu(dx)$, then the asymptotic variance increases by

$$E_a[X_0^2]^{-1} \int_{|x|<\epsilon} x^2 \, \mu(dx),$$

i.e. the jumps lead to an additional variance that is proportional to the intensity of jumps. This result motivates the jump filtering approach in Chapter 5 and 6. In the last part on the Ornstein-Uhlenbeck model we investigate the discretization error. We consider the estimator

$$\breve{a}_{\Delta_n,T} = \frac{\sum_{i=0}^{n-1} X_{t_i} \Delta_i X^c}{\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i}$$

instead of $\hat{a}_T$. Note that this is still a pseudo estimator, since the increments of the continuous martingale part are usually not observed. We prove that if $\Delta_n = o(T_n^{-2})$ then $\breve{a}_{\Delta_n,T}$ converges to the same asymptotic distribution as $\hat{a}_T$. Theorem 4.2.12 shows finally that the discretization bias when $\Delta_n = \Delta$ is kept fixed and we let $T \to \infty$ is of the order $O(\Delta)$ and can be reduced to $O(\Delta^2)$ via a bias correction.

In Section 4.3 we derive an explicit maximum likelihood estimator for the Lévy-driven square root process and prove that it is consistent and asymptotically normal. Efficiency follows again by proving the LAN property for the underlying statistical experiment. Then we generalize in Section 4.4 the results from the Ornstein-Uhlenbeck model and the square root process to jump diffusions with affine drift parameter, i.e.

$$\delta(\theta, s, x) = g(s, x) + \theta f(s, x)$$

for known functions $f, g : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$. This class also leads to an explicit and strongly consistent estimator, which is asymptotically normal under ergodicity. The last Section 4.5 contains the likelihood estimator for the stochastic delay equations with linear point delay. We show that it is strongly consistent and asymptotically normal. For the sake of simplicity we restrict our attention here to one-dimensional equations and delay measures that are supported on two points, but this can be extended easily to the multi-dimensional case with a more complex dependence on the past.

Chapter 5 is devoted to the problem of estimating the drift from discrete, arbitrarily spaced observations $X_{t_1}, \ldots, X_{t_n}$ for $0 = t_1 < \ldots < t_n = T_n$ when X has finite intensity jumps. Arbitrarily spaced means here that we only require that $\Delta_n \to 0$ fast enough such that $T_n \Delta_n^{(1-2\beta)\wedge\frac{1}{2}} = o(1)$, where β is the threshold exponent in the jump filter. Under the assumption of finite intensity the jump component of L can always be written as a compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Z_i$$

where N is a Poisson process and the Z_i's are iid with distribution F. To control the number of small jumps we suppose that $F(-2\Delta_n^\beta, 2\Delta_n^\beta) = o(T_n^{-1})$ as $n \to \infty$. When F has a bounded Lebesgue density this condition means that $\Delta_n^\beta T_n = o(1)$.

In this setting we prove for the drift estimator (1.3) for the Ornstein-Uhlenbeck type process that under stationarity

$$T_n^{1/2}(\bar{a}_n - a) \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2}{E_a[X_0^2]}\right) \text{ as } n \to \infty$$

under P^a. This convergence together with the efficiency result in Section 4.2 implies then asymptotic efficiency of $\bar{a}_n$. For the proof we define the good sets A_n^i as the events that a small increment of X implies that no jump occurred in that increment and vice versa:

$$A_n^i = \left\{\omega \in \Omega : \mathbf{1}_{\{|\Delta_i X| \le \Delta_n^\beta\}}(\omega) = \mathbf{1}_{\{\Delta_i N = 0\}}(\omega)\right\}.$$

We show then that the joint probability of the good sets tends to one as $n \to \infty$ such that in the limit the jump filter is able to identify all jumps of L. Then we prove that the continuous part is asymptotically not affected by the thresholding. In the last section of Chapter 5 we apply the jump filtering approach to obtain an estimator for linear stochastic delay differential equation from high-frequency observations and prove that it attains the same asymptotic distribution as the the likelihood estimator based on a fully observed process.

Chapter 6 contains as the final result of this thesis the proof that the likelihood approach with jump filtering for jumps of infinite activity leads to an asymptotically normal and efficient estimator. We restrict our attention here to the Ornstein-Uhlenbeck model. As in Chapter 5 it is necessary to control the behavior of small jumps of L. In order to separate continuous and jump part in the limit we suppose that the Lévy measure of L exhibits an α-stable like behavior around zero, i.e. there exists an $\alpha \in (0,2)$ such that

$$\int_{-v}^{v} x^2 \; \mu(dx) = O(v^{2-\alpha}) \tag{1.4}$$

as $v \downarrow 0$. This condition is closely related to the Blumenthal-Getoor index of L, which would be the minimal $\alpha \in (0,2)$ such that (1.4) holds. For α-stable Lévy processes this means that the stability index α satisfies (1.4). The second assumption on the jumps of L is that the small jumps are symmetric in a neighborhood of zero. If then there exists $\beta \in (0,1/2)$ such that $T_n \Delta_n^{1-2\beta\wedge\frac{1}{2}} = o(1)$ we obtain that $\bar{a}_n$ attains the efficient asymptotic distribution $N\left(0, \sigma^2 E_a[X_0^2]^{-1}\right)$. For the proof we derive first that the Markov generator of a pure jump Lévy process on a smoothed version of the test function $f^t(x) = x^2 \mathbf{1}_{\{|x| \le t^\beta\}}$ behaves like $O(t^{1+\beta(2+\alpha)})$ around zero. Then we separate the problem into the continuous martingale part plus jumps of finite activity and the remaining small jumps. Convergence of the first component follows from the results in Chapter 5. Then we apply the bound for the Markov generator and use the Lévy-Itô decomposition to proof that the component of small jumps is negligible in the limit.

In Chapter 7 we discuss simulation results to assess the finite sample behavior of the drift estimators from Chapter 5 and 6. First we consider models with finite jump activity and compare mean and standard deviation of Monte Carlo simulations for different jump intensities and parameter values. We also contrast the finite sample distribution with the asymptotic distribution from the central limit theorems. In the second part we perform a similar program for models with infinite jump activity. Overall we find that the estimators perform well if the maximal distance between observation is small enough such that the assumption of high-frequency observations is reasonable.

In the last Section 7.3 we compare the likelihood and the least squares estimator for the Ornstein Uhlenbeck model. It is well known that in the Gaussian case both estimators coincide. This is not the case anymore in models with jumps and as a corollary of Theorem 4.2.10 we obtain that the least squares estimator is inefficient in the jump case. This observation is confirmed also for finite samples, where we find that the likelihood estimator clearly outperforms the least squares approach when jumps are present and this performance gap becomes even larger with increasing jump intensity.

2 Basic theory and notation

The field of statistics of stochastic processes is a branch of stochastics that connects several modern parts of probability theory and mathematical statistics. In this chapter we fix our notation and collect in the first four sections the foundations from semimartingale theory with a special emphasis on local characteristics, Lévy processes and stochastic differential equations. In the second part in Section 2.5 and 2.6 we introduce with Le Cam's theory on asymptotics of statistical experiments and exponential families of stochastic processes two important concepts from modern statistics that will play a major role in this work.

2.1 Semimartingales

Semimartingales form a general class of stochastic processes that offers a rich theory of stochastic calculus. In the following we will fix our notation and collect some results based on Jacod and Shiryaev [2003]. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space.

Definition 2.1.1. An adapted, càdlàg process $Y : R_+ \times \Omega \to R^d$ is a *semimartingale* if a decomposition

$$Y = Y_0 + M + A$$

exists where $Y_0 \in \mathcal{F}_0$, M is a local martingale and A is a process of locally finite variation such that $M_0 = A_0 = 0$.

In the theory of semimartingales the quadratic (co-)variation process plays a central role. For the definition we need the notion of convergence in ucp. A sequence of processes X^n converges *uniformly on compacts in probability* (ucp) to a process X if for all $t \in \mathbb{R}_+$,

$$\sup_{0\leq s\leq t} \{|X_s^n - X_s|\} \xrightarrow{p} 0$$

as $n \to \infty$. Moreover, a sequence of stopping times $(\tau_n)_{n\in\mathbb{N}}$ is called a *random partition* if $\tau_0 = 0$ and $\sup_n\{\tau_n\} < \infty$ as well as $\tau_n < \tau_{n+1}$ for all n on the event $\{\tau_n < \infty\}$. A sequence of random partitions $(\tau_n^m)_{n,m\in\mathbb{N}}$ is called a *Riemann sequence* if

$$\sup_n\{|\tau_{n+1}^m \wedge t - \tau_n^m \wedge t|\} \to 0$$

as $m \to \infty$ for all $t \in \mathbb{R}_+$.

Theorem 2.1.2. *Let X, Y be two semimartingales. Then there exists a unique increasing, adapted, càdlàg process $[X, Y]$ such that for every Riemann sequence $(\tau_n^m)_{n,m\in\mathbb{N}}$ of random par-*

titions

$$\left(\sum_{i=1}^{\infty}(X_{\tau_{i+1}^m\wedge t}-X_{\tau_i^m\wedge t})(Y_{\tau_{i+1}^m\wedge t}-Y_{\tau_i^m\wedge t})\right)_{t\geq 0}\xrightarrow{ucp}[X,Y].$$

The process $[X,Y]$ is called the quadratic covariation of X and Y.

A proof of this result can be found in Jacod and Shiryaev [2003], Section I.4. As an alternative definition the relation

$$[X,Y]=XY-X_0Y_0-X_-\cdot Y-Y_-\cdot X$$

can often be found in the literature. From a statistical point of view the definition in Theorem 2.1.2 is interesting, since it suggest a natural estimator for $[X,Y]$ which is called the realized quadratic covariation.

The quadratic variation process leads also to the following important L^p-bound for the supremum of a martingale.

Theorem 2.1.3 (Burkholder-Davis-Gundy inequality)**.** *Let M be a càdlàg martingale and $p\geq 1$. Then there exist constants $c_p, C_p>0$ that do not depend on M such that*

$$c_pE\left[[M,M]_t^{p/2}\right]^{1/p}\leq E\left[\left(\sup_{s\leq t}\{M_s\}\right)^p\right]^{1/p}\leq C_pE\left[[M,M]_t^{p/2}\right]^{1/p}.$$

The constants appearing here are universal in sense that they depend on p, but not on M or the underlying probability space. A proof can be found in Chp. VII, Theorem 92 in Dellacherie and Meyer [1980].

2.1.1 Random measures

Our aim in this thesis is the development of statistical method for models that involve jump processes. In order to have a convenient description of the jump behavior of a càdlàg process we will use the language of random measures.

Definition 2.1.4. A random measure is a mapping $\rho:\mathcal{B}(R_+)\times\mathcal{B}(\mathbb{R}^d)\times\Omega\to\mathbb{R}_+$ such that $\rho(\cdot,\cdot,\omega)$ is a measure for each $\omega\in\Omega$ and $\rho(\{0\},\mathbb{R}^d,\omega)=0$ for all $\omega\in\Omega$.

For the definition of a stochastic integral with respect to a random measure ρ we refer to II.1d in Jacod and Shiryaev [2003]. We denote by $W*\rho$ the integral of an integrable function c with respect to ρ. The *optional σ-field* $\mathcal{O}$ on $\Omega\times\mathbb{R}_+$ is generated by all adapted càdlàg processes on $\Omega\times\mathbb{R}$. The σ-field $\mathcal{P}$ on $\Omega\times\mathbb{R}_+$ that is generated by all left-continuous processes is called the *predictable σ-field*. The following definition gives a suitable notion of measurability for random measures.

Definition 2.1.5.
(i) A random measure ρ is called *optional* if the process $W*\rho$ is $\mathcal{O}$-measurable for every $\mathcal{O}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable function W.

(ii) An optional random measure ρ is *σ-finite* if there exists a $\mathcal{P}\otimes\mathcal{B}(\mathbb{R}^d)$-measurable $V:\mathbb{R}_+\times\mathbb{R}^d\times\Omega\to(0,\infty)$ such that $\lim_{t\uparrow\infty}(V*\rho)_t$ is integrable (note that $V*\rho$ has a terminal variable, since V is strictly positive).

Now we are able to define the compensator of an optional σ-finite random measure ρ in full generality. By Theorem II.1.8 in Jacod and Shiryaev [2003] there exists a unique predictable random measure ν such that

$$E(W * \nu_\infty) = E(W * \rho_\infty)$$

holds for every nonnegative measurable function $W : \mathbb{R}_+ \times \mathbb{R}^d \times \Omega \to \mathbb{R}$.

Definition 2.1.6. The predictable random measure ν is called the compensator of ρ.

For us the most important example of a random measure is the jump measure of a càdlàg process.

Example 2.1.7. Let X be an adapted càdlàg process taking values in $\mathbb{R}^d$ and set $X_{t-} = \lim_{s\uparrow t} X_s$ and $\Delta X_s = X_t - X_{t-}$. Then

$$\rho(dt, dx, \omega) = \sum_s \mathbf{1}_{\{\Delta X_s(\omega)\neq 0\}} \delta_{(s,\Delta X_s(\omega))}(dt, dx) \tag{2.1}$$

defines a random measure with the following properties.

(i) $\rho(\{t\}, \mathbb{R}^d, \omega) \in \{0, 1\}$ for every ω, t,

(ii) ρ takes values in $\mathbb{N} \cup \{0, \infty\}$,

(iii) ρ is optional and σ-finite.

A random measure that has the properties (i) to (iii) is called an *integer-valued random measure*. When a càdlàg process has independent increments its associated random measure is a so-called Poisson random measure as defined below.

Definition 2.1.8. A *Poisson random measure* on $\mathbb{R}_+ \times \mathbb{R}^d$ is an integer-valued random measure ρ such that for $A \in \mathcal{B}(R_+) \times \mathcal{B}(\mathbb{R}^d)$ it holds that

(i) the measure defined by $\nu(A) = E[\rho(A)]$ is σ-finite and satisfies $\nu(\{t\} \times \mathbb{R}^d) = 0$ for $t \in \mathbb{R}_+$,

(ii) for every $t \in \mathbb{R}_+$ and if $A \subset (t, \infty) \times \mathbb{R}^d$ such that $\nu(A) < \infty$ then $\rho(\cdot, A)$ is independent of $\mathcal{F}_t$.

The compensator of a Poisson random measure is deterministic and given by $\nu(A) = E[\rho(A)]$ for $A \in \mathcal{B}(R_+) \times \mathcal{B}(\mathbb{R}^d)$. This is a consequence of the independence property (ii).

2.1.2 Semimartingale characteristics

The characteristics of a semimartingale are an extension of the Lévy-Khintchine triplet that describes the uniquely the law of a process with stationary and independent increments to semimartingales. They are a very useful tool in several different directions. For limit theorems the convergence of semimartingales can often be described by convergence of their characteristics (cf. Jacod and Shiryaev [2003]). In statistics the likelihood function of a semimartingale model can be given in terms of the characteristics as we will explore later on. They also form the basis for the formulation of martingale problems, solutions to absolute continuity problems and changes of measure in a semimartingale setting as we will see in Section 2.3. An nicely written introduction to semimartingale characteristics and their use in financial modeling can be found in Kallsen [2006].

As the Lévy-Khintchine triplet the characteristics of a semimartingale consist of three components that describe the generalized drift, a local martingale part and the jump behavior. In the following we collect the necessary notation for their definition.

Let $Y : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d$ be a semimartingale and $h : \mathbb{R}^d \to \mathbb{R}^d$ a truncation function, i.e. h is bounded, measurable and satisfies $h(x) = x$ in a neighborhood of 0. Let ΔY denote the adapted process defined by $\Delta Y_t = Y_t - Y_{t-}$ for $t \in \mathbb{R}_+$. We define the process of big jumps of Y by

$$\overline{Y}^h_t = \sum_{s \leq t} (\Delta Y_s - h(\Delta Y_s)) \tag{2.2}$$

and Y without its big jumps

$$Y^h = Y - \overline{Y}^h. \tag{2.3}$$

We have $\Delta Y^h = h(\Delta Y)$, such that Y^h has bounded jumps and therefore it admits a canonical decomposition (Jacod and Shiryaev [2003], Lemma I.4.24)

$$Y^h = Y(0) + M^h + B^h$$

where M^h is a local martingale with $M(0) = 0$ and B^h a predictable process of finite variation.

The jump characteristics stems from the jump measure of a semimartingale Y which is an integer-valued random measure $\rho : \Omega \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{N}$ defined by

$$\rho(\omega, dt, dx) = \sum_s 1_{\{\Delta Y_s(\omega) \neq 0\}} \delta_{(s, \Delta Y_s(\omega))}(dt, dx) \tag{2.4}$$

where $\delta_{(x,y)}$ denotes the Dirac measure with unit mass at (x, y).

Finally, every semimartingale $Y = Y_0 + M + A$ is by definition the sum of its starting value Y_0, a local martingale M and a process A of finite variation. By Theorem I.4.18 in Jacod and Shiryaev [2003] M exhibits a unique up to indistinguishablility decomposition $M = M^c + M^d$ into a continuous local martingale M^c and a purely discontinuous local martingale M^d. Recall that a local martingale M^d is purely discontinuous if for every continuous local martingale N the product $M^d N$ is a local martingale which means that M^d is orthogonal to the space of continuous martingales when square integrability holds. The uniqueness of this decomposition enables us to make the following definition.

Definition 2.1.9. The continuous local martingale M^c is called the *continuous martingale part* of Y and is denoted by $Y^c = M^c$.

Now we have collected all necessary notions to introduce semimartingale characteristics.

Definition 2.1.10. The *characteristic* of a semimartingale Y is the predictable triplet (B, C, ν), where $B = B^h$, $C = \langle Y^c, Y^c \rangle$ is the quadratic variation process of Y^c and ν is the predictable compensator of the jump measure ρ of Y.

Semimartingales were originally developed as the most general class of stochastic processes that allow for stochastic integration. This property can also be used as an alternative definition as was done in Protter [2004] to develop an alternative approach to semimartingale stochastic calculus. Since the class of semimartingales is closed under stochastic integration, it is an immediate question how the characteristics are transformed by the integral. In the following we use standard notation for stochastic integrals from Jacod and Shiryaev [2003] where $X \cdot Y = \int X \, dY$ denotes the stochastic integral of X with respect to Y and $X * \nu = \int X \, d\nu$ is the integral of X with respect to a random measure ν.

Proposition 2.1.11. *Let X be a d-dimensional semimartingale with characteristics (B, C, ν) relative to the truncation function h and H an $n \times d$-dimensional predictable process that is integrable with respect to X. Then the characteristics of $Y = \int H \, dX$ relative to the truncation function h' are (B', C', ν') where*

$$\begin{aligned} B' &= H \cdot B + (h'(Hx) - Hh(x)) * \nu, \\ C' &= \left(\sum_{k,l \le d} (H^{i,k} H^{j,l}) \cdot C^{k,l} \right)_{1 \le i,j \le n} \\ \nu'(A) &= \mathbf{1}_A(Hx) * \nu \textit{ for all } A \in \mathcal{B}(\mathbb{R}^n). \end{aligned}$$

2.2 Lévy processes

An important subclass of the class of semimartingales are Lévy processes. They provide a good compromise between the flexibility to fit to many kinds of dynamics in applications and their analytical tractability. In this section we will collect some basic results for Lévy processes.

2.2.1 Definition and characterization

We assume that a complete probability space $(\Omega, \mathcal{F}, P)$ is given and that it is equipped with a filtration $(\mathcal{F}_t)_{0 \le t < \infty}$. We say that a stochastic process $X = (X_t)_{0 \le t < \infty}$ is *continuous in probability* if for every $t \in \mathbb{R}_+$ and $\epsilon > 0$,

$$X_s \xrightarrow{p} X_t \text{ as } s \to t.$$

Definition 2.2.1. An adapted process $L = (L_t)_{0 \le t < \infty}$ is called a *Lévy process* if

1. $L_0 = 0$ almost surely,
2. L has *independent increments*, i.e. for all $n \in \mathbb{N}$ and $0 \le t_0 < \ldots < t_n$ the random variables $L_{t_0}, L_{t_2} - L_{t_1}, \ldots, L_{t_n} - L_{t_{n-1}}$ are independent.
3. L has *stationary increments*, i.e. for every $h \in \mathbb{R}_+$ the distribution of $L_{t+h} - L_t$ does not depend on t.
4. L is continuous in probability.

Two processes X and Y on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ are called *modifications* of each other if

$$P(X_t = Y_t) = 1 \text{ for all } t \in \mathbb{R}_+.$$

A proof of the following result can be found in Protter [2004], Theorem 30.

Theorem 2.2.2. *Every Lévy process has a modification with càdlàg paths.*

Therefore, we will assume throughout this thesis that we are working on the unique càdlàg version of any given Lévy process.

One reason why Lévy processes are popular is that this class of processes contains a large variety of different jump processes that make Lévy processes a versatile tool for stochastic modeling. These include classical examples like the Poisson or compound Poisson process, but also (tempered) stable, gamma and Normal-Inverse-Gaussian processes to name just a few of them.

For a given Lévy process L the behavior of its jumps can be conveniently described by its *Lévy measure*. It follows from the càdlàg property that every Lévy process has only finitely many jumps with jump sizes bounded away from zero. Hence, we can make the following definition.

Definition 2.2.3. Let L be a Lévy process. For every Borel set $B \subset \mathbb{R}\backslash\{0\}$ such that 0 is not in the closure of B, the measure defined by

$$\mu(B) = E\left[\sum_{0<s\leq 1} \mathbf{1}_B(L_s - L_{s-})\right]$$

is called the *Lévy measure* of L.

Remark 2.2.4. Every Lévy measure is finite on compacts except for a possible singularity around the origin. This singularity is such that

$$\mu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}} (x^2 \wedge 1)\, \mu(dx) < \infty. \tag{2.5}$$

The fundamental theorem in the theory of Lévy processes is an explicit factorization of the characteristic function for every Lévy process into a term that stems from a Gaussian process, a deterministic drift component and a term that characterizes the jump behavior. This decomposition is the famous Lévy -Khinthchine formula. It was first derived in special cases by de Finetti and Kolmogorov. Later Paul Lévy proved the general cased for $\mathbb{R}^d$ valued Lévy processes. A much simpler proof for the one-dimensional version was given by Khintchine. A proof can be found in Bertoin [1998] or Sato [1999].

Theorem 2.2.5 (Lévy-Khintchine formula)**.** *Let L be a Lévy process. Then there exists a characteristic triplet (b, σ^2, μ) consisting of $b \in \mathbb{R}$, $\sigma^2 \in \mathbb{R}_+$ and a Lévy measure μ on $\mathbb{R}$ such that*

$$\phi_{L_t}(u) = E\left[e^{iuL_t}\right] = e^{t\psi(u)} \tag{2.6}$$

where

$$\psi(u) = ibu - \frac{\sigma^2}{2}u^2 + \int_{\mathbb{R}} \left(e^{iux} - 1 - iux\mathbf{1}_{\{|x|\leq 1\}}\right) \mu(dx).$$

Conversely, for every triplet (b, σ^2, μ) *as above such that* μ *satisfies the integrability condition* (2.5) *there exists a Lévy process* L *with characteristic function of the form* (2.6)*.*

The characteristic triplet is also called the *Lévy-Khintchine triplet.* Every Lévy process is a semimartingale and vice versa every semimartingale with deterministic and constant local characteristics is a Lévy process and its Lévy-Khintchine triplet (b, σ^2, μ) and semimartingale characteristics (B, C, ν) are then related by

$$\begin{aligned} B(\omega, t) &= bt, \\ C(\omega, t) &= \sigma^2 t, \\ \nu(\omega, dt, dx) &= \mu(dx) \otimes \lambda(dt), \end{aligned}$$

where λ denotes the Lebesgue measure on $\mathbb{R}$. A proof of this result is given in II.4.19 of Jacod and Shiryaev [2003].

2.2.2 Distributional and path properties

The Lévy -Khintchine formula shows that the law of a Lévy process is uniquely determined by the characteristic triplet (b, σ^2, μ). We notice immediately that the characteristic function of a Lévy process factorizes into the characteristic function of a Brownian motion with drift $\phi_{W_t+tb}(u) = \exp(ibu - \frac{\sigma^2}{2}u^2)$ and the integral with respect to μ. A corresponding decomposition of L exists also in a path-wise sense. This representation is the so-called *Lévy -Itô decomposition.* For a detailed proof we refer to Sato [1999] or Jacod and Shiryaev [2003].

Theorem 2.2.6 (Lévy -Itô decomposition)**.** *For every Lévy process* L *with characteristic triplet* (b, σ^2, μ) *there exist a Wiener process* W *and a Poisson random measure*

$$N : \mathcal{B}(\mathbb{R}_+) \times \mathcal{B}(\mathbb{R}) \times \Omega \to \mathbb{N}$$

with compensator μ *such that* W *and* N *are independent and*

$$\begin{aligned} L_t &= W_t + bt + \int_0^t \int_{|x|<1} x \left(N(dt, dx) - dt\mu(dx)\right) + \int_0^t \int_{|x|\leq 1} x\, N(dt, dx) \\ &= W_t + bt + \int_0^t \int_{|x|<1} x \left(N(dt, dx) - dt\mu(dx)\right) + \sum_{s\leq t} \Delta L_s \mathbf{1}_{\{|\Delta L_s|\leq 1\}}. \end{aligned}$$

When in addition to (2.5) we have $\mu(\mathbb{R}) < \infty$ then we say that jumps of L are of *finite activity.* This corresponds to the case that the jump part of L is of compound Poisson type. In this case the Lévy -Itô decomposition reads as follows:

$$L_t = W_t + bt + \sum_{s\leq t} \Delta L_s.$$

The following formulae from Kunita [2010] for the first two conditional moments of integrals with respect to a Poisson random measure N can be easily proved by considering the usual extension argument from simple functions to square integrable functions.

Proposition 2.2.7. *Let f be an $\mathcal{F}_s \times \mathcal{B}(\mathbb{R})$-measurable random variable such that*

$$E\left[\int_{\mathbb{R}\backslash\{0\}} f(x)^2\ \mu(dx)\right] < \infty.$$

Then for any $s < t$ a.s.,

$$E\left[\int_s^t \int_{\mathbb{R}\backslash\{0\}} f(x)\ (N(dt,dx) - dt\mu(dx)) \middle| \mathcal{F}_s\right] = 0, \tag{2.7}$$

$$E\left[\left(\int_s^t \int_{\mathbb{R}\backslash\{0\}} f(x)\ (N(dt,dx) - dt\mu(dx))\right)^2 \middle| \mathcal{F}_s\right] = (t-s)\int_{\mathbb{R}\backslash\{0\}} f(x)^2\ \mu(dx). \tag{2.8}$$

2.3 Some tools from stochastic analysis

Stochastic analysis provides many important ideas for the statistical analysis of stochastic processes. In the first section we introduce the concept of martingale problems in a semimartingale setting. They will prove useful together with Hellinger processes to solve absolute continuity problems and develop the likelihood theory for jump diffusions in Chapter 3. The last part of this section collects some martingale limit theorems that will be needed for the asymptotic analysis of our estimators.

2.3.1 Martingale problems

Martingale problems were originally developed in the study of diffusion processes to understand the relation between the coefficients $a : \mathbb{R}_+ \times \mathbb{R}^d \to R^{d\times d}$ and $b : \mathbb{R}_+ \times \mathbb{R}^d \to R^d$ of a diffusion process X and the distribution $P_{s,x}$ of X starting in $x \in \mathbb{R}^d$ at time $s \in \mathbb{R}_+$. When the generator G of X is given by

$$G_t = \frac{1}{2}\sum_{i,j=1}^d a_{ij}(t,\cdot)\frac{\partial^2}{\partial_i\partial_j} + \sum_{i=1}^d b_i(t,\cdot)\frac{\partial}{\partial_i},$$

then for all $f \in C_0^\infty(\mathbb{R}^d)$ and $s \in \mathbb{R}_+$ fixed an application of Itô's formula shows that the process

$$f(X_t) - \int_s^t G_u f(X_u)\, du, \quad t \geq s \tag{2.9}$$

is a $P_{s,x}$-martingale with respect to a given filtration $(\mathcal{G}_t)_{t\geq 0}$ such that $\mathcal{F}_t^X \subseteq \mathcal{G}_t$ for all $t \geq 0$. This is the first example of a martingale problem. In other words a martingale problem asks for existence and uniqueness (in a suitable sense to be defined later) of measures $P_{s,x}$ under which (2.9) is a martingale and the initial condition

$$P_{s,x}(X_s = x) = 1 \tag{2.10}$$

holds. More material on martingale problems in the setting of diffusion processes can be found in the book by Stroock and Varadhan [2006]. This idea was than generalized to the setting of semimartingales where the coefficients a, b were replaced by a set of semimartingale characteristics (B, C, ν) and one asks for the existence and uniqueness of a measure such that a given process is a semimartingale under this measure with characteristics (B, C, ν). In the following we will mainly follow Chapter III in Jacod and Shiryaev [2003]. An even more comprehensive source on martingale problems in a semimartingale setting is Jacod [1979].

Let (B, C, ν) be a predictable triplet on a filtered space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ relative to the truncation function $h : \mathbb{R} \to \mathbb{R}$ such that:

1. $B : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is an $(\mathcal{F}_t)$-predictable process with finite variation and $B(0) = 0$;
2. $C : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is an $(\mathcal{F}_t)$-predictable continuous and nonnegative process with $C(0) = 0$;
3. the random measure $\nu : \Omega \times \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}$ is $(\mathcal{F}_t)$-predictable such that $\nu(\mathbb{R}_+ \times \{0\}) = \nu(\{0\} \times \mathbb{R}) = 0$ and $(|x|^2 \wedge 1) * \nu_t(\omega) < \infty$ as well as $\int h(x)\nu(\omega; \{t\} \times dx) = \Delta B_t(\omega)$ and $\nu(\omega; \{t\} \times \mathbb{R}) \leq 1$ for all $\omega \in \Omega$ and all $t \geq 0$.

Suppose now that Y is a càdlàg process on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$. We formalize the problem of finding a measure P on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ under which Y is a semimartingale with characteristics (B, C, ν) as follows.

Definition 2.3.1. A measure P on $(\Omega, \mathcal{F})$ solves the martingale problem associated with Y, initial distribution π on $(\Omega, \mathcal{F}_0)$ and (B, C, ν) if:

(i) Under P the distribution of $Y(0)$ equals π.

(ii) Y is semimartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ with characteristics (B, C, ν) relative to h.

We denote by $s(Y|\pi; B, C, \nu)$ the set of all solution measures P of the martingale problem associated with the process Y, initial distribution π and characteristics (B, C, ν). If the initial distribution is clear from the context, we will often drop it.

Besides uniqueness in the sense that $\#s(Y|\pi; B, C, \nu) = 1$ we will also need the following notion of uniqueness. Let $(\mathcal{F}_t^0)$ denote the filtration generated by Y.

Definition 2.3.2. (i) A mapping $T : \Omega \to \bar{\mathbb{R}}_+$ such that $\{T \leq t\} \in \mathcal{F}_t^0$ for all $t \in \mathbb{R}_+$ is called a strict stopping time.

(ii) Local uniqueness holds for a martingale problem $s(Y|\pi; B, C, \nu)$ if for every strict stopping time T any two solutions $P, P' \in s(Y^T|\pi; B^T, C^T, \nu^T)$ of the stopped problem coincide on $\mathcal{F}_T^0$.

All martingale problems that we will encounter in this work will be related to solutions of stochastic differential equations. Of course there is a close connection between martingale problems and stochastic differential equations. The next theorem is one result in this direction that relates weak solutions of an SDE to the solutions of a corresponding martingale problem.

Let W be an m-dimensional Wiener process and μ a Poisson random measure with compensator $\nu = \lambda \otimes F$ for a σ-finite measure F on $\mathbb{R}^d$ both defined on some filtered space $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t\geq 0}, P')$. Consider the following stochastic differential equation driven by W and μ.

$$\begin{aligned} dY_t &= \beta(t, Y_t)\, dt + \gamma(t, Y_t)\, dW_t + h \circ \delta(t, Y_{t-}, z)(\mu(dt, dz) - \lambda(dt) \otimes F(dx)) \\ &\quad + (x - h(x)) \circ \delta(t, Y_{t-}, z)\mu(dt, dz) \\ Y_0 &= \xi \end{aligned} \tag{2.11}$$

where h is a truncation function, the initial value ξ is $\mathcal{F}'_0$-measurable and the coefficients are Borel measurable mappings

$$\begin{aligned} \beta &: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d, \\ \gamma &: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times m}, \\ \delta &: \mathbb{R}_+ \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d. \end{aligned}$$

Theorem 2.3.3. *Let $\mathcal{P}(\beta, \gamma, \mu, \xi)$ be the set of all distributions of weak solutions of* (2.11) *on the canonical space $(\Omega, \mathcal{F})$. Then*

$$\mathcal{P}(\beta, \gamma, \mu, \xi) = s(X|\pi; B, C, \nu)$$

where $\pi = \mathcal{L}(\xi)$ and‘

$$\begin{aligned} B_t &= \int_0^t \beta(s, X_s)\, ds, \\ C_t &= \int_0^t \gamma\gamma\top(s, X_s)\, ds, \\ \nu(\omega, dt, dx) &= \lambda(dt) \otimes K_t(X_t, dx), \\ K_t(x, A) &= \int \mathbf{1}_{A\setminus\{0\}}(\delta(t, y, z))\, F(dz). \end{aligned} \tag{2.12}$$

The next theorem is an existence and uniqueness result for martingale problems related to jump diffusion processes that will be useful when we investigate absolute continuity questions for solutions of stochastic differential equations.

Theorem 2.3.4. *Assume that (B, C, ν) are as in* (2.11) *such that β is bounded, $\gamma\gamma^\top$ is bounded continuous and everywhere invertible and*

$$(t, y) \mapsto \int_A (|z|^2 \wedge 1) K_t(y, dz)$$

is bounded and continuous for every $A \subset \mathcal{B}(\mathbb{R}^d)$. Then $s(X|\pi; B, C, \nu)$ has a unique solution.

2.3.2 Absolute continuity and singularity of measures

Given a measure P on a filtered measurable space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ we denote by P_t the restriction of P to $\mathcal{F}_t$.

Definition 2.3.5. We say that P' is locally absolutely continuous with respect to P, if for the restrictions $P'_t \ll P_t$ holds for every $t \in \mathbb{R}_+$. We will write $P \overset{loc}{\ll} P$ in this case.

Let P and P' denote two probability measures on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ and $h(\alpha, P, P'), \alpha \in [0,1)$ any version of the Hellinger process as defined in Jacod and Shiryaev [2003], IV.1.

Theorem 2.3.6. *Let T be a stopping time and assume $\mathcal{F} = \mathcal{F}_{\infty-}$. Then the following are equivalent*

(i) $P'_T \ll P_T$,

(ii) $P'_0 \ll P_0$ *and* $P'(h(\alpha)_T < \infty) = 1$ *for any* $\alpha \in (0,1)$ *and* $P'(h(0)_T = 0) = 1$,

(iii) $P'_0 \ll P_0$ *and* $h(\alpha)_T \to 0$ *under* P' *as* $\alpha \downarrow 0$.

A proof can be found in Jacod and Shiryaev [2003], IV.2.

2.3.3 Limit theorems

Here we will collect several limit theorems for (local) martingales that we will use frequently in this thesis. Let $(\Omega, \mathcal{F}, P)$ together with $(\mathcal{F}_t)_{t\leq 0}$ be a filtered probability space.

Laws of large numbers

To prove consistency of various estimators we will need the following law of large numbers from Liptser [1980].

Theorem 2.3.7. *Let M be a locally square integrable martingale with $M_0 = 0$ and suppose that A is a predictable, non-decreasing and right-continuous process with $A_0 = 0$. Define*

$$B_t = \int_0^t (1 + A_s)^{-2} \, d\langle M\rangle_s,$$

then as $t \to \infty$,

$$\frac{M_t}{A_t} \xrightarrow{a.s.} 0 \text{ on } \{A_\infty = \infty\} \cap \{B_\infty < \infty\}.$$

When we take the deterministic function $A_s = s$ for A in Theorem 2.3.7 we obtain as an immediate corollary a law of large numbers for locally square integrable martingales normalized by the quadratic variation process.

Theorem 2.3.8. *Let M be locally square integrable martingale. Then as $t \to \infty$ we have*

$$\frac{M_t}{\langle M\rangle_t} \xrightarrow{a.s.} 0$$

on $\{\langle M\rangle_\infty = \infty\}$.

Central limit theorems

In this section we give a central limit theorem for multivariate martingales. It will be important later on to investigate the relative error of the estimators under consideration. A proof can be found in Küchler and Sørensen [1999]. For a more detailed discussion on martingale limit theorems the read may consult Hall and Heyde [1980] or Jacod and Shiryaev [2003].

Theorem 2.3.9. *Let M be an n-dimensional square integrable martingale with mean zero and covariance matrix $H_t = E[M_t^2]$. Assume that there exists an invertible non-random sequence $(k_t)_{t\geq 0}$ of $n \times n$-matrices such that $k_t \to 0$ as $t \to \infty$ and*

1. *for $K_t^i = \sum_{j=1}^n |k_{jit}|$ we have*
$$k_t E\left[\sup_{s\leq t} |\Delta M_s|\right] \to 0 \text{ as } t \to \infty,$$
2. $$k_t[M_t]k_t^\top \xrightarrow{p} W \text{ as } t \to \infty$$
where W is a random positive semi-definite matrix such that $P(\det(W) > 0) > 0$,
3. $$k_t H_t k_t^\top \to \Sigma \text{ as } t \to \infty$$
and $\Sigma \in \mathbb{R}^{n\times n}$ is positive definite.

Then as $t \to \infty$
$$(k_t M_t, k_t[M_t]k_t^\top) \xrightarrow{\mathcal{D}} (W^{1/2}Z, W)$$
and conditionally on $\{\det(W) > 0\}$
$$W^{-1/2}k_t M_t \xrightarrow{\mathcal{D}} Z,$$
where Z is n-dimensional standard normal distributed independent of W.

2.4 Stochastic differential equations

Since the introduction of Itô's stochastic integral, stochastic differential equations have become a central topic in modern probability theory. Their solutions form prototypic examples for Markov processes and semimartingales whose generator and semimartingale characteristics can be explicitly given in terms of the coefficients of the equation itself. Therefore very powerful tools are at hand for their study.

In applications SDEs provide an intuitive approach for stochastic modeling. They form a natural extensions of models that use ordinary differential equations to situations where measurement error or other sources of uncertainty are present. Hence, they are nowadays standard tools in fields like mathematical finance and economics, physics, engineering or biology.

2.4.1 Basic notions and results

We will formulate the basic results in a general semimartingale setting. For a more specialized theory that is framed for Lévy processes we recommend Applebaum [2009]. We will mainly follow Jacod [1979] and Protter [2004]. For our purpose it will be sufficient to consider strong solutions under Lipschitz conditions on the coefficients. Assume that a filtered probability space $(\Omega, \mathcal{F}, P)$ and $(\mathcal{F}_s)_{s\leq 0}$ is given.

Definition 2.4.1. strong solution and uniqueness

Next we give a standard existence and uniqueness result under Lipschitz conditions.

Theorem 2.4.2. *Let Y be a semimartingale starting at zero and $f : \mathbb{R} \times \mathbb{R}_+ \times \Omega \to \mathbb{R}$ be such that*

1. *For every $x \in \mathbb{R}$ the random process $f(x) : \mathbb{R}_+ \times \Omega \to \mathbb{R}$ is $\mathcal{F}_t$-adapted and has càdlàg paths.*
2. *Lipschitz condition on f: there exists a finite random variable K such that*

$$|f(x,t,\omega) - f(y,t,\omega)| \leq K(\omega)|x-y|$$

for all $(x, y, t, \omega) \in \mathbb{R}^2 \times \mathbb{R}_+ \times \Omega$.

If in addition $X_0 \in \mathcal{F}_0$ is finite then

$$X_t = X_0 + \int_0^t f(X_{s-}, s, \cdot)\, dY_s$$

has a unique strong solution that is also a semimartingale.

Example 2.4.3 (Lévy-driven Ornstein-Uhlenbeck processes). Let $(L_t, t \geq 0)$ be a Lévy process on a given filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t), P)$ adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$. We call for every $a \in \mathbb{R}$ a strong solution X to the stochastic differential equation

$$dX_t = -aX_t\, dt + dL_t, \quad t \in \mathbb{R}_+, \quad X_0 = \tilde{X}, \tag{2.13}$$

an Ornstein Uhlenbeck (OU) process driven by the Lévy process L with initial distribution $\pi = \mathcal{L}(\tilde{X})$. The initial condition $\tilde{X}$ is assumed to be independent of L.

It follows from Itô's formula that an explicit solution of (2.13) is given by

$$X_t = e^{-at}X_0 + \int_0^t e^{-a(t-s)} dL_s, \quad t \in \mathbb{R}_+. \tag{2.14}$$

The integral in (2.14) can by partial integration be defined path-wise as a Riemann-Stieltjes integral , since the integrand is of finite variation (see Dudley [2002] for example). This solution to equation (2.13) is unique up to indistinguishability.

Equation 2.13 admits a stationary solution (cf. Sato and Yamazato [1984]) if and only if

$$\int_{|x|>1} \log|x|\; \mu(dx) < \infty \text{ and } a > 0. \tag{2.15}$$

Under these conditions X has a unique invariant distribution F and $X_t \xrightarrow{\mathcal{D}} X_\infty \sim F$ as $t \to \infty$. The Lévy-Itô decomposition 2.2.6 for L can be used to relate properties of X and L. As an example the following lemma provides the second moment of the stationary distribution of X in terms of the Lévy-Khintchine triplet of L.

Lemma 2.4.4. *Assume that* (2.15) *holds and denote the Lévy-Khintchine triplet of L by* (b, σ^2, μ). *Then*

$$E_a\left[X_\infty^2\right] = b_\infty^2 + \frac{\sigma^2}{2a} + \int_{\mathbb{R}} x^2 \, \mu_\infty(dx),$$

where

$$\begin{aligned} b_\infty &= \frac{b}{a} + \int_{\mathbb{R}} \int_0^\infty e^{-as} z \left(\mathbf{1}_{(-1,1)}(e^{-as} z) - \mathbf{1}_{(-1,1)}(z) \right) \, ds \, \mu(dz), \\ \mu_\infty(A) &= \int_0^\infty \mu(e^{-as} A) \, ds, \quad A \in \mathcal{B}(\mathbb{R}). \end{aligned}$$

The statement of the lemma remains valid when L has infinite second moment such that also the second moment of X_∞ is infinite.

Proof: It was shown in Theorem 4.1 and 4.2 in Sato and Yamazato [1984] that under (2.15) the process X has a unique invariant distribution that is self-decomposable and exhibits the Lévy -Khintchine triplet $(b_\infty, \sigma^2/2a, \mu_\infty)$. Hence, the characteristic function of X_∞ is given by

$$\phi_{X_\infty}(u) = \exp\left(iub_\infty - \frac{1}{4a} u^2 \sigma^2 + \int_{\mathbb{R}} (e^{iuz} - 1 - iuz\mathbf{1}_{(-1,1)}(z)) \, \mu_\infty(dz) \right).$$

The second moment of X_∞ results in

$$E_a[X_\infty^2] = -\phi''_{X_\infty}(0) = b_\infty^2 + \frac{\sigma^2}{2a} + \int_{\mathbb{R}} x^2 \, \mu_\infty(dx).$$

□

Since every Lévy process is a semimartingale and the class of semimartingales is closed under transformations such as (2.14), X is again a semimartingale. Next, we will derive its characteristics.

Lemma 2.4.5. *Let* X *be an Ornstein-Uhlenbeck process driven by a Lévy process L with characteristic triplet* (b, σ^2, ρ) *then the semimartingale characteristics* (B, C, ν) *of* X *are given by*

$$\begin{aligned} B(\omega, t) &= bt - a \int_0^t X_s(\omega) ds, \\ C(\omega, t) &= \sigma^2 t, \\ \nu(\omega, dt, dx) &= \rho(dx) \, \lambda(dt), \end{aligned}$$

where λ *denotes the Lebesgue measure on* $\mathbb{R}$.

Proof: Since the characteristics of a semimartingale do not depend on the initial distribution, we set without loss of generality $X_0 = 0$. If we write the OU equation (2.13) in integral form

$$X_t = -a \int_0^t X_s ds + L_t,$$

B and C follow from Proposition IX.5.3 in Jacod and Shiryaev [2003] and the fact that the semimartingale characteristics of L are $(bt, \sigma^2 t, \rho(dx)\lambda(dt))$.

By I.4.36 in Jacod and Shiryaev [2003] it follows from equation (2.14) that X and L have the same jump measure. Hence, the compensator of their jump measures coincide. □

2.4.2 Stochastic delay differential equations

Solutions to stochastic differential equations whose coefficients depend also on past values of the process are particular examples of Itô or diffusion type processes (cf. Liptser and Shiryaev [2001]). Even if they can still be represented as solutions to stochastic differential equations the dependence on the past of the process destroys the Markov property of the solutions. An interesting example are solutions X to stochastic delay differential equations of the form

$$\begin{aligned} dX_t &= \int_{-r}^0 X(t+u)\, m(du)\, dt + dL_t, \quad t > 0, \\ X_t &= X_t^0 \quad t \in [-r, 0], \end{aligned} \tag{2.16}$$

where m is a finite signed measure on a finite interval $[-r, 0]$, L is a Lévy process with Lévy-Khintchine triplet (b, σ^2, μ) and $(X_t^0)_{-r \le t \le 0}$ is an initial process with càdlàg paths. Since no stochastic integrals are involved in defining (2.16), solutions can be understood in a path-wise sense. If X^0 is integrable a unique path-wise solutions exists. The existence of a stationary solution was investigated in Mohammed and Scheutzow [1990] when L is a general semimartingale and by Gushchin and Küchler [2000] for Lévy driver. In the following give a brief summary of their results.

The fundamental solution of the deterministic equation corresponding to (2.16) is a function $x_0 : [-r, \infty) \to \mathbb{R}$ such that

$$\begin{aligned} x_0(t) &= \int_{-r}^0 x_0(t+u)\, a(du), \quad t > 0, \\ x_0(t) &= 0, \quad t \in [-r, 0), \\ x_0(0) &= 1. \end{aligned}$$

The solution of (2.16) in terms of x_0 is then

$$\begin{aligned} X_t &= x_0(t) X_0^0 + \int_{-r}^0 \int_u^0 X_s^0 x_0(t+u-s)\, ds\, a(du) + \int_0^t L_{t-s}\, dx_0(s), \quad t \ge 0, \\ X_t &= X_t^0, \quad t \in [-r, 0). \end{aligned}$$

We see now that the asymptotic behavior of X is determined by the asymptotic behavior of

x_0. In turn the asymptotic behavior of x_0 is characterized by the solutions of the so-called characteristic equation

$$h(\lambda) = \lambda - \int_{-r}^{0} e^{\lambda u}\, a(du) = 0.$$

It follows now from Lemma 2.1 in Gushchin and Küchler [2000] that $x_0(t) = O(e^{v_0 t})$ where

$$v_0 = \sup\{\mathrm{Re}\lambda | h(\lambda) = 0\}.$$

This relation between the fundamental solution and X can be used to give necessary and sufficient conditions for the existence of a stationary solution.

Theorem 2.4.6 (Theorem 3.1, Gushchin and Küchler [2000]). *Equation* (2.16) *admits a stationary solution if and only if one of the following equivalent conditions hold.*

(i) $v_0 < 0$ *and*

$$\int_{|x|>1} \log(|x|)\, \mu(dx) < \infty.$$

(ii) There exists a solution X *such that* X_t *converges in distribution as* $t \to \infty$.

(iii) Any solution X *has a limit distribution* $X_t \xrightarrow{\mathcal{D}} X_\infty$ *as* $t \to \infty$.

These results will prove useful in Chapter 4 when we derive limiting results for drift estimators for linear stochastic delay equations.

2.5 Le Cam theory

The foundations of modern statistical decision theory and the theory of statistical experiments was laid by Blackwell and Wald. Blackwell [1951] introduced the notion of statistical experiments as the triplet

$$(\Omega, \mathcal{F}, (P^\theta)_{\theta\in\Theta})$$

were $(\Omega, \mathcal{F})$ is a measurable space and $(P^\theta)_{\theta\in\Theta}$ a family of probability measures on $(\Omega, \mathcal{F})$. The main idea in the theory of statistical experiments is to compare different experiments by defining a suitable distance and related mode of convergence for them. First ideas in this direction go back to Wald [1943].

In the second half of the 20th century Lucien Le Cam developed these ideas further into his theory on asymptotics of statistical experiments via localization. When we consider a sequence of experiments $(P_n^\theta)_{\theta\in\Theta}^{n\in\mathbb{N}}$ then typically for two different θ, θ' the sequence of measures (P_n^θ) and $(P_n^{\theta'})$ will be singular in the limit, i.e. we obtain a trivial limit experiment. To obtain a meaningful limit experiment the idea of localization was introduced which means a rescaling of the parametrization in analogy to the normalization in the central limit theorem. Le Cam [1960] found that the prototypic limit experiment is the so-called Gaussian shift experiment which takes the role of the Gaussian law in the central limit theorem. An experiment is then called locally asymptotically normal (LAN) when its likelihood converges to the likelihood of the Gaussian shift experiment. When the LAN property is fulfilled very powerful results such as the minimax theorem by Hájek and Le Cam or Hájek's convolution theorem are at hand. But of course many other limiting behaviors have been found so far. In more complex models based on stochastic processes local asymptotic mixed normality (LAMN) and local asymptotic quadraticity (LAQ) often appear. A quite intricate example in the context of stochastic delay differential equations was investigated in Gushchin and Küchler [1999]. For a modern viewpoint on the convergence of statistical experiments in the general framework of (γ, Γ)-models and λ-convergence the reader may consult Shiryaev and Spokoiny [2000].

In the following we will be working on a filtered measurable space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$. That together with a family $(P^\theta)_{\theta\in\Theta}$ of probability measures forms our statistical experiment. We assume that $\Theta \subset \mathbb{R}^d$ and that a dominating probability measure P exists such that $P_t^\theta \ll P_t$ for all $\theta \in \Theta$ and $t \in \mathbb{R}_+$. Here P_t denotes the restriction of P to the σ-field $\mathcal{F}_t$. The log-likelihood function will be denoted by $l_t(\theta) = \log(dP_t^\theta / dP_t)$.

Definition 2.5.1. The statistical experiment $(P^\theta)_{\theta\in\Theta}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ is called

(i) *locally asymptotically quadratic* at θ if there exist a d-dimensional random process $V_t(\theta)$ and a sequence of symmetric positive semidefinite random and adapted $k \times k$-matrices $I_t(\theta)$ such that $I_t(\theta)$ is P^θ-a.s. positive definite and

$$l_t(\theta + \delta_t h) - l_t(\theta) - h^\top I_t(\theta)^{1/2} V_t(\theta) - \frac{1}{2} h^\top I_t(\theta) h = o_p(1)$$

for every $h \in \mathbb{R}^d$ and every sequence of positive definite matrices $\delta_t \in \mathbb{R}^{d\times d}$ such that $\delta_t \overset{t\to\infty}{\longrightarrow} 0$. The sequence $(V_t(\theta), I_t(\theta))_{t\geq 0}$ is bounded in probability as $t \to \infty$, the sequences $(P_t^{\theta+\delta_t h})_{t\geq 0}$ and $(P_t^\theta)_{t\geq 0}$ are contiguous and if $(V_\infty(\theta), I_\infty(\theta))$ is a cluster point of $(V_t(\theta), I_t(\theta))_{t\geq 0}$ then $I_\infty(\theta)$ is a.s. positive definite.

(ii) *locally asymptotically mixed normal* at θ if (i) holds and

$$(V_t(\theta), I_t(\theta)) \xrightarrow{\mathcal{D}} (V(\theta), I(\theta)) \text{ as } t \to \infty$$

under P_θ where $I(\theta)$ is a P_θ-a.s. positive definite random matrix such that conditionally on $I(\theta)$ the random vector $V(\theta)$ is standard normal.

(iii) *locally asymptotically normal* at θ if (ii) holds such that $I(\theta)$ is deterministic.

When the LAMN property holds it can be shown that the matrix $I(\theta)$ is a lower bound for the asymptotic variance of an estimator for θ in the sense of the following theorem. Let $\mathring{\Theta}$ denote the interior of Θ.

Theorem 2.5.2 (convolution theorem)**.** *Let $(P^\theta)_{\theta\in\Theta}$ be LAMN at $\theta \in \mathring{\Theta}$ and suppose that $(\hat\theta_t)_{t\geq 0}$ is a sequence of estimators such that $\theta_t \in \mathcal{F}_t$ and*

$$(I_t(\theta), \delta_t^{-1}(\hat\theta_t - \theta - \delta_t h)) \xrightarrow{\mathcal{D}} (I(\theta), \Sigma(\theta))$$

under $P_t^{\theta+\delta_t h}$. *Then there exists a kernel* $k(I(\theta))$ *such that almost surely*

$$P(\Sigma(\theta)|I(\theta)) = k(I(\theta)) * N(0, I(\theta)^{-1}).$$

The next theorem gives a lower bound for the risk of an estimator of θ under arbitrary loss functions.

Theorem 2.5.3 (minimax theorem). *Let* $(P^\theta)_{\theta\in\Theta}$ *be LAMN at any* $\theta \in \mathring{\Theta}$ *and suppose that* $\hat{\theta}_t$ *is any estimator for* θ. *Then for any bowl-shaped loss function* $f : \mathbb{R}^d \to [0, \infty)$ *we have*

$$\lim_{r\to\infty} \liminf_{t\to\infty} \sup_{|\delta_t^{-1}(\theta'-\theta)|\leq r} E_{\theta'} \left[f(\delta_t^{-1}(\hat{\theta}_t - \theta')) \right] \geq E\left[f(I(\theta)^{-1}V(\theta) \right] = E\left[f(I(\theta)^{-1/2}Z(\theta) \right]$$

where $Z \sim N(0, Id_d)$ *is independent of* I.

We will use these results to define optimality of an estimator in the following sense.

Definition 2.5.4. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\leq 0}, (P^\theta)_{\theta\in\Theta})$ be as statistical experiment. An estimator $\hat{\theta}_t$ of θ is called *asymptotically efficient* if its (conditional) asymptotic variance attains the lower bound given in Theorem 2.5.2.

2.6 Exponential families

Classical exponential families of probability distribution have a long history in mathematical statistics. On one hand they provide a suitably general class of models for many applications including important examples like normal, Poisson, exponential distribution and at the same time these models are quite easy to handle from a mathematical point of view such that in many cases optimal inference methods are at hand.

This idea of a general class of statistical models that have many desirable properties and include the major part of models that we use in applications was then generalized to the theory of exponential families of stochastic processes. Since also the models that we are going to investigate in this thesis lead to exponential families of stochastic processes, many results of this theory will be of great use to us. In this section we give a suitable definition for our purpose and collect some important results closely following the book by Küchler and Sørensen [1997].

Let $\{P^\theta, \theta \in \Theta\}$ be a family of measures on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0})$ with parameter set $\Theta \subset \mathbb{R}^d$ such that the interior $\mathring{\Theta}$ is nonempty. We assume that a dominating measure P^{θ_0} for some $\theta_0 \in \Theta$ exists such that local absolute continuity $P^\theta \overset{loc}{\ll} P^{\theta_0}$ holds (recall Definition 2.3.5) for all $\theta \in \Theta$.

Definition 2.6.1. A statistical experiment $\{P^\theta, \theta \in \Theta\}$ forms a *curved exponential family* if the likelihood function exists and is of the form

$$\mathcal{L}_t(\theta, \omega) = \frac{dP_t^\theta}{dP_t^{\theta_0}} = \exp\left(\theta^\top A_t - \kappa(\theta) S_t\right) \tag{2.17}$$

where $\kappa : \Theta \to \mathbb{R}$ and $A : \Omega \times \mathbb{R}_+ \to \mathbb{R}^d$ is a càdlàg process. Moreover, $S : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ is assumed to be a non-decreasing continuous process with $S_0 = 0$ and $S_t \overset{t\to\infty}{\longrightarrow} \infty$ P^θ-a.s. for all $\theta \in \Theta$.

Let us take a look at some examples.

Example 2.6.2. Consider the Lévy-driven Ornstein-Uhlenbeck process X from Example 2.4.3. Then X is a solution to the stochastic differential equation

$$dX_t = -\theta X_t \, dt + dL_t, \quad t \in \mathbb{R}_+, \quad X_0 = \tilde{X},$$

where L is a Lévy process and $\theta \in \mathbb{R}$. When a is our parameter of interest such that $\Theta = \mathbb{R}$, then the class of Ornstein-Uhlenbeck processes index by a induces a family of measures $\{P^\theta, \theta \in \mathbb{R}\}$ on the space of càdlàg functions on $[0, \infty)$. We will prove in Chapter 3 that under suitable conditions on L the family $\{P^\theta, \theta \in \mathbb{R}\}$ forms a curved exponential family with likelihood function

$$\mathcal{L}_t(\theta, X^T) = \exp\left(-\frac{\theta}{\sigma^2}\int_0^T X_{s-} \, dX_s^c - \frac{\theta^2}{2\sigma^2}\int_0^T X_s^2 \, ds\right).$$

So for

$$\begin{aligned} A_t &= \frac{1}{\sigma^2}\int_0^t X_s \, dX_s^c \\ \kappa &= \frac{\theta^2}{2\sigma^2} \\ S_t &= \int_0^t X_s^2 \, ds \end{aligned}$$

we see that the class of Lévy-driven Ornstein-Uhlenbeck processes forms a curved exponential family.

Example 2.6.3. Another example is the class of Cox-Ingersoll-Ross process, which can be defined as a solutions of

$$dX_t = -\theta X_t dt + \sqrt{X_t} dL_t; \quad 0 \le t \le T$$

driven by a Lévy process $L_t = W_t + bt + J_t$, where W is a Wiener process, $b > 0$ and J is assumed to be subordinator (see Sato [1999], Section 4.21 on subordinators). The starting value is $X_0 = x > 0$ and the drift parameter $\theta > 0$. These conditions assure that the process stays positive such that the square-root is well-defined. In this example the likelihood function (cf. Chapter 3) is

$$\mathcal{L}_t(\theta, X) = \exp\left(-\frac{\theta}{\sigma^2}(X_T^c - X_0^c) - \frac{\theta^2}{2\sigma^2}\int_0^T X_t \, dt\right). \tag{2.18}$$

which is again of the form of an exponential family as in (2.17).

In the setting of an exponential family the maximum likelihood estimator is consistent and asymptotically normal under weak regularity conditions given in the following two results. They are taken from Küchler and Sørensen [1997], Section 5.2. A strictly convex function $\kappa : \Theta \to \mathbb{R}$

is called *steep* if for all $\theta_1 \in \Theta \backslash \mathring{\Theta}$, $\theta_0 \in \mathring{\Theta}$ and $\theta_s = (1-s)\theta_0 + s\theta_1$, $s \in (0,1)$,

$$\frac{d}{ds}\kappa(\theta_s) \to \infty \quad \text{as } s \uparrow 1.$$

Theorem 2.6.4. *Suppose a curved exponential family of the form* (2.17) *is given such that* κ *is steep and* $S_t > 0$ *for* $t > 0$ P^θ*-a.s. for all* $\theta \in \Theta$*. Then the maximum likelihood estimator* $\hat{\theta}_T$ *from time continuous observations on* $[0,T]$ *exists and is uniquely given by*

$$\hat{\theta}_T = (\dot{\kappa})^{-1}\left(\frac{A_T}{S_T}\right)$$

if and only if $A_T/S_T \in \mathring{C}$ *where* $C = \dot{\kappa}(\mathring{\Theta})$.

If also $\theta \in \mathring{\Theta}$*, then under* P^θ *the maximum likelihood estimator exists and is unique for* T *sufficiently large and* $\hat{\theta}_T \xrightarrow{a.s.} \theta$ *as* $T \to \infty$.

Next we give the corresponding central limit theorem.

Theorem 2.6.5. *For a curved exponential family of the form* (2.17) *assume that* $\theta \in \mathring{\Theta}$ *and that there exists an increasing positive non-random function* $\phi_\theta(t)$ *such that under* P^θ

$$\phi_\theta(t)^{-1} S_t \xrightarrow{p} \eta(\theta)^2$$

as $t \to \infty$ *where* $\eta(\theta)^2$ *is a finite non-negative random variable for which* $P^\theta(\eta(\theta)^2 > 0) > 0$*. Then under* P^θ

$$\left(S_t^{-1/2}\left(A_t - \kappa(\theta)S_t\right), \phi_\theta(t)^{-1}S_t\right) \xrightarrow{\mathcal{D}} N\left(0, \ddot{\kappa}(\theta)\right) \times F_\theta$$

as $t \to \infty$ *conditionally on* $\{\eta(\theta)^2 > 0\}$*, where* F_θ *is the conditional distribution of* $\eta(\theta)^2$ *given* $\{\eta(\theta)^2 > 0\}$*. Moreover,*

$$(S_t^{1/2}(\hat{\theta}_t - \theta), \phi_\theta(t)) \xrightarrow{\mathcal{D}} N(0, \ddot{\kappa}(\theta)^{-1}) \times F_\theta$$

conditionally on $\{\eta(\theta)^2 > 0\}$ *as* $t \to \infty$.

3 Likelihood theory for jump diffusions

In this chapter we give existence results for the likelihood function of solutions to stochastic differential equations driven by Lévy processes when continuous observations are given. This leads naturally to absolute continuity problems for the measures induced by these processes on the path space. Conditions for absolute continuity can be given in terms of the Hellinger process that describes roughly speaking the time evolution of Hellinger integrals corresponding to different solutions. In turn the Hellinger processes can then be given in terms of the semimartingale characteristics of two processes.

When the measures on the path space are equivalent the Radon-Nikodym derivative is known explicitly. In the drift estimation problem considered here the density process does not depend on the jump part of the processes, in the sense that the jump measures do not appear in the density process. At the same time the density process involves explicit knowledge of the continuous martingale part. Since the continuous martingale part is not directly observed, this will be our main challenge in the second part of this work when only discrete observations are assumed.

The general theory of absolute continuity and singularity of measures in a semimartingale setting was developed by Jacod [1979]. We will tailor these results to the setting of jump diffusions and discuss several examples in the spirit of Sørensen [1991]. Likelihood theory for ergodic diffusion processes driven by Brownian motion was treated in the book by Kutoyants [2004].

3.1 Jump diffusion processes

Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space and L be an m-dimensional Lévy process on this space with Lévy -Khintchine triplet $(b, \Sigma\Sigma^\top, \mu)$. Define X as a solution to

$$X_t = X_0 + \int_0^t \delta(\theta, s, X_s)\, ds + \int_0^t \gamma(s, X_{s-})\, dL_s, \quad t \in \mathbb{R}_+, \tag{3.1}$$

where $X_0 \in \mathbb{R}^d$, the parameter set $\Theta \subset \mathbb{R}^n$ is such that $\mathring{\Theta} \neq \emptyset$ and the coefficients are Borel measurable functions

$$\begin{aligned} \delta &: \Theta \times \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d, \\ \gamma &: \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^{d\times m}, \end{aligned}$$

such that γ takes values in the positive definite matrices and the following regularity conditions are satisfied.

Assumption 3.1.1. 1. Local Lipschitz condition: There exists $C_n \in \mathbb{R}_+$ such that for all $t, |x|, |y| \leq n$ and $\theta \in \Theta$,

$$\begin{aligned} |\delta(\theta, t, x) - \delta(\theta, t, y)| &\leq C_n |x - y|, \\ |\gamma(t, x) - \gamma(t, y)| &\leq C_n |x - y|. \end{aligned}$$

2. Linear growth condition: For each $n \in \mathbb{N}$ there exists $C_n \in \mathbb{R}_+$ such that for all $t \leq n$, $x \in \mathbb{R}^d$ and $\theta \in \Theta$:

$$\begin{aligned} |\delta(\theta, t, x)| &\leq C_n(1 + |x|), \\ |\gamma(t, x)| &\leq C_n(1 + |x|). \end{aligned}$$

Under these conditions Theorem 2.4.2 guarantees existence and uniqueness of a strong solution to equation (3.1). The solution X is a semimartingale and hence has càdlàg paths. As a mapping into the path space X induces a measure P^θ on the canonical space $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t \geq 0})$ (cf. Jacod and Shiryaev [2003], III.2.13 for the canonical setting). Then under P^θ the canonical process is again a semimartingale with characteristics

$$\begin{aligned} B_t(\theta, \omega) &= \int_0^t \beta(\theta, s, X_s(\omega)) \, ds, \\ C_t(\omega) &= \int_0^t c(s, X_s(\omega)) \, ds, \\ \nu(\omega, dy, dt) &= K_t(X_{t-}(\omega), dy) \otimes \lambda(dt), \end{aligned} \tag{3.2}$$

where

$$\begin{aligned} \beta(\theta, s, x) &= \delta(\theta, s, x) + b + \int_{|y|>1} y \; K_s(x, dy), \\ c(s, x) &= \gamma \Sigma \Sigma^\top \gamma^\top (s, x), \\ K_s(x, A) &= \int_{\mathbb{R}^m} \mathbf{1}_{A \setminus \{0\}}(\gamma(s, x) y) \; \mu(dy), \quad \forall A \in \mathcal{B}(\mathbb{R}^d). \end{aligned} \tag{3.3}$$

By abuse of notation we denote the canonical process likewise by X. In the following we assume time continuous observations of the canonical process $(X_t)_{0 \leq t \leq T}$ over $[0, T]$. The next result gives conditions for absolute continuity of the measures induced by X for different parameters and an explicit representation of the Radon-Nikodym derivative for the case where absolute continuity holds.

Theorem 3.1.2. *Suppose that Assumption 3.1.1 holds, c is strictly positive definite and*

$$(t, x) \mapsto \int_A (|y|^2 \wedge 1) \; K_t(x, dy)$$

is continuous for all $A \in \mathcal{B}(\mathbb{R}^d)$. *Define for* $\theta, \theta' \in \Theta$ *the process*

$$h_t(\theta,\theta') = \int_0^t a(\theta,\theta',s,X_s)^\top c(s,X_s)^{-1} a(\theta,\theta',s,X_s)\, ds,$$

where

$$a(\theta,\theta',s,x) = \beta(\theta,s,X) - \beta(\theta',s,x).$$

Then

(i) $P_t^\theta \overset{loc}{\ll} P_t^{\theta'}$ *if and only if*

$$P^\theta(h_t(\theta,\theta') < \infty) = P^{\theta'}(h_t(\theta,\theta') < \infty) = 1.$$

(ii) *If (ii) holds the likelihood function is given by*

$$\frac{dP_t^\theta}{dP_t^{\theta'}} = \exp\Bigg[\int_0^t c(s,X_{s-})^{-1} a(\theta,\theta',s,X_{s-})\, dX_s^c \\ -\frac{1}{2}\int_0^t a(\theta,\theta',s,X_s)^\top c(s,X_{s-})^{-1} a(\theta,\theta',s,X_s)\, ds\Bigg]$$

here X^c *denotes the continuous martingale part under* $P^{\theta'}$.

The theorem follows from Theorem 2.1 in Sørensen [1991], which can be proved by applying Theorem III.5.34 in Jacod and Shiryaev [2003] by establishing local uniqueness for the corresponding martingale problem. We have given here a simplified version for parametric estimation of the drift. In the next sections we discuss in detail examples where the maximum likelihood estimator can be derived explicitly.

3.2 Lévy-driven Ornstein-Uhlenbeck processes

As a first example for a jump diffusion model that is frequently used in applications let us recall the stochastic differential equation that defines the one-dimensional Lévy-driven Ornstein-Uhlenbeck process (cf. Section 2.4.1).

$$dX_t = -aX_t\, dt + dL_t, \quad t \in \mathbb{R}_+, \quad X_0 = \tilde{X}. \tag{3.4}$$

Here L is a one-dimensional Lévy process with Lévy-Khintchine triplet (b, σ^2, μ), $\tilde{X}$ is a possibly random initial condition and $a \in \mathbb{R}$ the unknown drift parameter. Denote by P^a and $P^{a'}$ the measures induced by X on the path space for two different drift parameters $a, a' \in \mathbb{R}$. To illustrate the results from the previous section we provide here conditions in terms of the Lévy-Khintchine triplet of L for absolute continuity and derive the Radon-Nikodym derivative. In order to simplify our notation we will write from now on P', B' and so on for all objects that correspond to a'.

Theorem 3.2.1. *Let $P^a, P^{a'}$ be two solution measures of the OU equation for the same driving Lévy process L with characteristic triplet (b, σ^2, ρ) and initial distributions π and π'. Suppose $\sigma^2 > 0$, $\pi' \ll \pi$ and μ has a finite second moment, then we have $P^{a'} \overset{loc}{\ll} P^a$.*

Before we proceed to the proof let us give a short example when absolute continuity fails.

Example 3.2.2. Suppose the driving Lévy process is of compound Poisson form

$$L_t = \sum_{i=1}^{N_t} Y_i,$$

where N is a Poisson process and the Y_i's are iid random variables. In this case obtain that the induced measures $\{P^a, a \in \mathbb{R}\}$ are mutually singular, since the trajectories of X corresponding to the parameter a are piecewise exponentials with rate a such that the induced measures have disjoint support.

We divide the proof into three steps. First, define a martingale problem that gives weak solutions to equation (3.4). Then, we will derive the Hellinger process of two different solution measures and finally the Hellinger process will solve the absolute continuity/singularity problem. Some basic facts on martingale problems can be found in Section 2.3.1.

Now we derive the Hellinger process for two solution measures of equation (3.4). For the corresponding theory on Hellinger distance and absolute continuity and singularity problems we refer the reader to Jacod and Shiryaev [2003], IV.1.

Proposition 3.2.3. *A version of the Hellinger process H corresponding to solution measures $P^a, P^{a'}$ of the Ornstein-Uhlenbeck equation* (2.13) *is given by*

$$H_t(\alpha, a, a') = (a' - a)^2 \frac{\alpha(1-\alpha)}{2\sigma^2} \int_0^t X_s^2 ds, \quad t \in \mathbb{R}_+,$$

where $\alpha \in (0, 1)$.

Proof: Denote by (B, C, ν) the characteristics of X with initial distribution π as in Lemma 2.4.5. The problem of finding a solution measure to the defining SDE (2.13) is equivalent to the martingale problem $s(X|\pi; B, C, \nu)$, since by Theorem III.2.26 in Jacod and Shiryaev [2003] the solution measures to the SDE are given by $s(X|\pi; B, C, \nu)$ (by abuse of notation we use the same expression for the problem and the set of solutions).

Next, we define a predictable process by

$$\tilde{B}_t = B_t - B'_t = (a' - a) \int_0^t X_s ds.$$

Set $\tilde{\beta}(a, a')_s = \frac{a'-a}{\sigma^2} X_s$. Then according to IV.3.9 in Jacod and Shiryaev [2003] a candidate for the Hellinger process corresponding to two solutions of the martingale problems $s(X|\pi; B, C, \nu)$ and $s(X|\pi'; B', C, \nu)$ is given by

$$H_t(\alpha, a, a') = \frac{\alpha(1-\alpha)\sigma^2}{2} \int_0^t \tilde{\beta}(a, a')_s^2 1_\Sigma \, ds. \tag{3.5}$$

In our setting we obtain

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

Hence, $H_t(\alpha, a, a')$ takes the form

$$H_t(\alpha, a, a') = (a' - a)^2 \frac{\alpha(1-\alpha)}{2\sigma^2} \int_0^t X_s^2 ds, \quad t \in \mathbb{R}_+.$$

The next step is to show that $H_t(\alpha, a, a')$ is indeed a version of the Hellinger process. This follows from local uniqueness of the martingale problems $s(X|\pi; B, C, \nu)$ and $s(X|\pi'; B', C, \nu)$, which follow from Corollary III.2.41 in the case of jump diffusions. Hence, by Corollary IV.3.68 in Jacod and Shiryaev [2003] $h(\alpha, a, a')$ is indeed a version of the Hellinger process. $\square$

The relation $P_t \ll P_t'$ is by Theorem 2.3.6 equivalent to $H_t(\alpha, a, a') < \infty$ a.s. under P' such that Theorem 3.2.1 follows now easily by bounding H on $(0, t)$.

Proof of Theorem 3.2.1: We have shown that a version of the Hellinger process of X under P and P' is given by $H_t(\alpha, a, a')$ from Proposition 3.2.3 and in our setting both processes have the same initial distribution. Hence, in order to prove that $P' \overset{loc}{\ll} P$ it is sufficient by Theorem IV.2.1 in Jacod and Shiryaev [2003] to show that $P'(H_t(\frac{1}{2}, a, a') < \infty) = 1$. Thus, we need that

$$\int_0^t \left[\int_0^{u-} \left(e^{-a'(u-s)} - e^{-a(u-s)} \right) L(ds) \right]^2 du < \infty, \quad P_{a'}\text{-a.s. } \forall t \in [0, T].$$

Set $f(u, s) = e^{-a'(u-s)} - e^{-a(u-s)}$. By the Lévy-Itô decomposition we can write L as

$$L_t = \sigma W_t + J_t + bt + \sum_{s \le t} \Delta L_s \mathbf{1}_{\{|\Delta L_s| \ge 1\}},$$

where W is a Wiener process and $J_t = \int_{\{|x|<1\}} x(N_t(dx) - t\mu(dx))$ with jump measure N of L and its compensator μ. Define $M_t = \sigma W_t + J_t$ and $V_t = \sum_{s \le t} \Delta L_s \mathbf{1}_{\{|\Delta L_s| \ge 1\}} + bt$. Then

$$E_a \left[\left(\int_0^u f(u, s)\, dL_s \right)^2 \right] \le E_a \left[\left(\int_0^u f(u, s)\, dM_s \right)^2 \right] + E_a \left[\left(\int_0^u f(u, s)\, dV_s \right)^2 \right]$$

Since V is of finite variation, it can be decomposed path-wise into the difference of two increasing functions such that $E_a \left[(\int_0^u f(u, s)\, dV_s)^2 \right] < \infty$ follows by bounding f by its supremum and using that V is locally bounded.

M is a martingale such that we can apply Burkholder-Davis-Gundy inequality to obtain

$$E_a \left[\left(\int_0^u f(u, s)\, dM_s \right)^2 \right] \le C E_a \left[\int_0^u f(u, s)^2\, d[M]_s \right].$$

Since $[M]$ is increasing and $0 \leq f(u,s) \leq 1$ within the range of integration, it follows that

$$E_a\left[\int_0^u f(u,s)^2\, d[M]_s\right] \leq E\left[[M]_u\right] = \sigma^2 u + E_a\left[\sum_{s\leq u}(\Delta L_s)^2 \mathbf{1}_{\{|\Delta L_s|\leq 1\}}\right]$$
$$= \sigma^2 u + u\int_0^1 x^2\, \mu(dx) < \infty$$

again by the Lévy-Itô decomposition and Theorem 2.3.8 in Applebaum [2009]. □

When absolute continuity holds, the Radon-Nikodym density can be given explicitly. Recall the definition of the continuous martingale part of a semimartingale from Section 2.1.

Proposition 3.2.4. *If $\sigma^2 > 0$ and $\pi \ll \pi'$, then the Radon-Nikodym density process of P' with respect to P is given by*

$$Z_t = \frac{dP'_t}{dP_t} = \frac{dP'_0}{dP_0}\exp\left(\frac{(a-a')}{\sigma^2}\int_0^t X_s\, dX_s^c - \frac{(a-a')^2}{2\sigma^2}\int_0^t X_s^2 ds\right) \tag{3.6}$$

P-a.s. and X^c denotes the continuous martingale part of X under P.

Proof: If (B, C, ν) and (B', C, ν) are the characteristics of X under P and P', respectively, it follows from Lemma 2.4.5 that we can write

$$B' = B + \sigma^2 \int_0^{\cdot} \beta dt,$$

where we have introduced a predictable process $\beta_t = \frac{(a-a')}{\sigma^2} X_t$.

Next, we define another random process by

$$K = \int_0^{\cdot} \sigma^2 \beta^2 dt = \frac{(a-a')^2}{\sigma^2}\int_0^{\cdot} X_t^2 dt.$$

K is predictable and has non-decreasing paths. Obviously, K does not jump to infinity in the sense that $K_{T-} = \infty$ on $\{T = \inf(t : K_t = \infty) < \infty\}$. We set

$$T_n = \inf(t \in \mathbb{R}_+ : K_t \geq n), \quad A = \bigcup_n [0, T_n].$$

T_n is a stopping time and A defines a predictable random set. Since

$$K_t = \frac{(a-a')^2}{\sigma^2}\int_0^t X_s^2 ds \xrightarrow{a.s.} \infty$$

under P and P' as $t \to \infty$, also $T_n \xrightarrow{a.s.} \infty$ under P and P' such that $A = \mathbb{R}$ a.s. under P and P'. We know already from the previous proposition that local uniqueness holds for the martingale problem $s(X|\pi; B', C, \nu)$. By III.5.10/32 in Jacod and Shiryaev [2003] there exists now a process

$U : A \to \mathbb{R}$ such that
$$U^S = \left(\frac{(a-a')}{\sigma^2} X_t 1_{[0,S]}\right) \cdot X^c,$$
for every stopping time S with $[0,S] \subset A$ and Z is given by
$$Z_t = Z_0 \exp\left(U_t - \frac{(a-a')^2}{2\sigma^2}\int_0^t X_s^2 ds\right) \prod_{s \le t}(1+\Delta U_s)e^{-\Delta U_s}, \qquad t \in [0,S].$$
Since U is continuous, the product term disappears.

Finally, we observe that the continuous martingale part X^c of X is a Wiener process under P^0. Indeed, by the Lévy-Itô decomposition of the driving Lévy process L we can write X as (cf. Protter [2004], Theorem I.42)
$$X_t = X_0 - a\int_0^t X_s\, ds + \sigma W_t + J_t \quad , t \ge 0, \tag{3.7}$$
where W is a Wiener process and J a quadratic pure jump process (cf. Protter [2004], p.71) given by
$$J_t = \int_{\{|x|<1\}} x(N_t(dx) - t\mu(dx)) + bt + \sum_{0 \le s \le t} \Delta X_s 1_{\{|\Delta X_s| \ge 1\}}, \tag{3.8}$$
here μ is the Lévy measure, b the drift of L and N denotes the Poisson random measure on $\mathbb{R}$ associated with the jumps of L that is independent of W. The integral term in (3.7) is of finite variation and N is purely discontinuous such that we have decomposed the semimartingale X into a process of finite variation
$$A_t = -a\int_0^t X_s\, ds + bt + \sum_{0 \le s \le t} \Delta X_s 1_{\{|\Delta X_s| \ge 1\}}$$
and a local P^0-martingale
$$M_t = \sigma W_t + \int_{\{|x|<1\}} x(N_t(dx) - t\mu(dx)).$$
By Theorem I.4.18 in Jacod and Shiryaev [2003] this decomposition of M is unique such that $X^c = \sigma W$ under P^0 up to indistinguishability. □

Remark 3.2.5. For the proof of local equivalence in Theorem 3.2.1 we have employed general results on absolute continuity problems for semimartingales (cf. Jacod [1979] and Jacod [1979]) that also apply for general jump diffusions as in Theorem 3.1.2.

For the Ornstein-Uhlenbeck model an alternative proof follows by defining the process
$$Z_t = \exp\left(\frac{a}{\sigma^2}\int_0^t X_s\, dX_s^c - \frac{a^2}{2\sigma^2}\int_0^t X_s^2 ds\right)$$

and showing that Z is a martingale with $E[Z_t] = 1$ that induces a new measure via

$$P_t^a = Z_t P_t^0 \quad \text{for all } t \in \mathbb{R}_+$$

such that X is an Ornstein-Uhlenbeck process under P^a. The last step follows by deriving the characteristics of X under P^a via a version of Girsanov's theorem (cf. Theorem III.39 in Protter [2004]) and using that X is the unique solution of the Langevin equation.

3.3 Square-root processes

This class of processes was first studied by William Feller in Feller [1951], who coined the name square-root processes. The square-root process has the remarkable property that due to the square-root term in the diffusion coefficient it stays nonnegative at all times. This property and its analytical tractability led to its popularity in mathematical finance, where it has been applied in interest rate modeling, derivative pricing and stochastic volatility models. In term structure modeling Cox et al. [1985] introduced this class of processes for the dynamics of the short rate. Whereas the Heston model employs a square-root process as a stochastic volatility process of a price process (cf. Heston [1993]). All these classical references consider square-root processes driven by a Wiener process only.

The square-root process with jumps X can be defined as a solution to the following stochastic differential equation

$$dX_t = -aX_t dt + \sigma\sqrt{X_t}dW_t + dL_t, \quad 0 \le t < \infty, \tag{3.9}$$

driven by a Lévy process $L_t = bt + J_t$ and a standard Wiener process W, where $\sigma, b > 0$ and J is assumed to be subordinator. When the starting value $X_0 = x > 0$ and drift parameter $a > 0$ are strictly positive the square root term secures that X stays nonnegative at all times.

Existence and uniqueness of a nonnegative strong solution of (3.9) have been proved in Section 5 in Dawson and Li [2006]. A model of this form with a jump component has been employed in Barndorff-Nielsen and Shephard [2001] to include leverage in their stochastic volatility model. Kallsen [2006] derives the semimartingale characteristics of X and shows that X has an affine structure such that X is also a regular affine Feller process (cf. Duffie et al. [2003] on regular affine processes). It follows from Corollary 3.2 in Mayerhofer et al. [2011] that X stays strictly positive for all $t \ge 0$ if $b \ge \sigma^2/2$.

The paths of X lie in $D[0,\infty)$ such that for every $a > 0$ the process induces a measure P^a on $D[0,\infty)$. Necessary and sufficient conditions for absolute continuity for two measures P^a and $P^{a'}$ corresponding to different coefficients follow easily from Theorem 3.1.2. We summarize these results in the following proposition which also provides the density process in the case when absolute continuity holds.

Proposition 3.3.1. *Suppose that L has a nonzero Gaussian component and that $P_0^a \ll P_0^0$.*

Then $P^a \overset{loc}{\ll} P^0$ *and the density process is given by*

$$\frac{dP_T^a}{dP_T^0} = \frac{dP_0^a}{dP_0^0} \exp\left(-\frac{a}{\sigma^2}(X_T^c - X_0^c) - \frac{a^2}{2\sigma^2}\int_0^T X_t\, dt\right) \tag{3.10}$$

P^0*-almost surely and* X^c *denotes the continuous* P^0*-martingale part.*

3.4 Stochastic delay differential equations

In this section we leave the class of jump diffusions that we have investigated so far to demonstrate that the likelihood approach also works in a non-Markovian setting for delay equations. We have already introduced some basic material on stochastic delay equations driven by Lévy processes in Section 2.4.2. We specialize the setting of Section 2.4.2 to linear delay equations in order to obtain a likelihood function that is a quadratic in the parameter. We assume that the measure a that describes the past dependence of the process is supported on finitely many points $0 = \tau_0 < \ldots < \tau_m < \infty$ for some $m \in \mathbb{N}$.

Let L be a Lévy process with Lévy-Khintchine triplet (b, σ^2, μ) on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ and X the solution to the stochastic delay differential equation

$$\begin{aligned} dX_t &= \sum_{i=0}^m a_i X_{t-\tau_i}\, dt + dL_t, \quad t \geq 0, \\ X_t &= X_t^0, \quad t \in [-\tau_m, 0). \end{aligned} \tag{3.11}$$

The initial condition X^0 is a càdlàg process and independent of L and X is parametrized by the drift parameter $a = (a_0, \ldots, a_m) \in \mathbb{R}^{m+1}$. For $a_1 = \ldots = a_n = 0$ we obtain the Ornstein-Uhlenbeck type process from Section 3.2.

For each $a \in \mathbb{R}^{m+1}$ the process X induces a measure P^a on the space of càdlàg functions $D[-\tau_m, \infty)$. If $\sigma^2 > 0$ these measures are mutually locally absolutely continuous.

$$P^a \overset{loc}{\ll} P^{a'} \quad \text{for all } a, a' \in \mathbb{R}^{m+1}.$$

This follows from (2.10) in Küchler and Sørensen [1989], since X is by Theorem V.3.7 in Protter [2004] a semimartingale. The Radon-Nikodym derivative takes the form

$$\frac{dP_T^a}{dP_T^0} = \exp\left(a^\top V_T - \frac{1}{2} a^\top I_T a\right) \tag{3.12}$$

for $T > 0$ where the $m+1$-dimensional process V equals

$$V_T = \left(\int_0^T X_{t-\tau_i}\, dX_t^c\right)_{i=0,\ldots,m},$$

where the continuous martingale part X^c is taken under the dominating measure P^0 and the

$(m+1) \times (m+1)$-dimensional process I is given by

$$I_T = \left(\int_0^T X_{t-\tau_i} X_{t-\tau_j} \, dt \right)_{i,j=0,\dots,m}.$$

The form of $\frac{dP_T^a}{dP_T^0}$ in 3.12 implies that the system of equations forms an exponential family in the sense of Section 2.6. In the next chapter we will use these results to construct a maximum likelihood estimator for a.

4 Maximum likelihood estimation

In this chapter we develop a maximum likelihood approach for estimating the drift of a jump diffusions from time-continuous observations and discuss examples of jump diffusion models that lead to an explicit maximum likelihood estimator. We prove asymptotic normality results and give conditions for efficiency of the maximum likelihood estimator in the sense of Hájek-Le Cam.

We also investigate how the activity of jumps influences the estimation error and find that the asymptotic variance is proportional to the intensity of jumps. This result motivates the jump filtering approach in the next chapters when the problem of estimation from discrete observations is considered and we approximate the continuous martingale part by using a threshold technique. As a consequence we obtain that the least squares estimator is inefficient in models with jumps.

For the Ornstein-Uhlenbeck process we discuss asymptotic properties of the discretized likelihood estimator. These results lay the foundation for the next two chapters, where we prove asymptotic properties for the discretized estimator under approximation of the continuous martingale part.

For a comparison to diffusions without jumps driven by Brownian motion we refer the reader to the book by Kutoyants [2004], where a general maximum likelihood theory for these models is developed.

It should be made clear at this point that all estimators in this chapter are only pseudo estimators in the sense that they are build on time-continuous observations, which are in general not available in practice. The problem of estimation from discrete observations will be considered in Chapter 5 and 6, where we will build upon the results developed here. We will use them in two directions. First of all we construct the discrete estimators from the estimators in this chapter and second the pseudo estimators will serve us as a benchmark when we come to efficiency questions for the discrete case.

4.1 General jump diffusion processes

Consider the jump diffusion X as a solution to

$$X_t = X_0 + \int_0^t \delta(\theta, s, X_s)\, ds + \int_0^t \gamma(s, X_{s-})\, dL_s, \quad t \in \mathbb{R}_+, \tag{4.1}$$

where $X_0 \in \mathbb{R}^d$ and L, δ and γ are as in Section 3.1 and the drift coefficient is parametrized by $\theta \in \Theta \subset \mathbb{R}^n$. We denote the Lévy -Khintchine triplet of L by $(b, \Sigma\Sigma^\top, \mu)$. Suppose that $\delta(0, \cdot, \cdot) \equiv 0$ and that time-continuous observations $(X_t)_{t\in[0,T]}$ are given. Define

$$c(s, X_s) = \gamma\Sigma\Sigma^\top\gamma(s, X_s)^\top$$

and denote by P^θ the measure induced by X on the path space $D[0,\infty)$. If c is invertible the likelihood function for θ is by Theorem 3.1.2 equal to

$$\mathcal{L}(X,\theta)_T = \exp\left(\int_0^T c(s,X_s)^{-1}\delta(\theta,s,X_s)\, dX_s^c \right.$$
$$\left. - \frac{1}{2}\int_0^T \delta(\theta,s,X_s)^\top c(s,X_s)^{-1}\delta(\theta,s,X_s)\, ds\right),$$

where X^c is the continuous martingale part under P^0. A maximum likelihood estimator for θ is then any $\hat{\theta}_T$ such that

$$\hat{\theta}_T = \arg\max_{\theta\in\Theta} \mathcal{L}(X,\theta)_T. \tag{4.2}$$

In many widely used diffusion models this equation can be solved explicitly. In the next two sections we discuss two examples of this kind, the Ornstein-Uhlenbeck type and the square-root or Cox-Ingersoll-Ross process, in detail. In the last section we move on to models that depend in an affine way on the parameter such that the likelihood equation (4.2) still exhibits an explicit solution.

4.2 Ornstein-Uhlenbeck type processes

As a first example for a jump diffusion model that leads to an explicit maximum likelihood estimator we consider the Lévy-driven Ornstein-Uhlenbeck process. Recall the stochastic Langevin equation (cf. Section 2.4.1).

$$dX_t = -aX_t\, dt + dL_t, \quad t\in\mathbb{R}_+, \quad X_0 = \tilde{X}. \tag{4.3}$$

Here L is a Lévy process with Lévy-Khintchine triplet (b,σ^2,μ), $\tilde{X}$ is a possibly random initial condition and $a\in\mathbb{R}$ the unknown drift parameter. The unique strong solution X to this equation will be called an Ornstein-Uhlenbeck process in the following.

From the results of Proposition 3.2.4 follows now an explicit maximum likelihood estimator $\hat{a}$ for the coefficient a of the OU process. By setting $\pi=\pi'$ in (3.6) we obtain that the likelihood function with respect to the dominating measure P^0 is

$$L(a;X^T) = \exp\left(-\frac{a}{\sigma^2}\int_0^T X_s\, dX_s^c - \frac{a^2}{2\sigma^2}\int_0^T X_s^2\, ds\right),$$

where X^c denotes the continuous P^0-martingale part of X. Maximizing the log-likelihood yields finally

$$\hat{a}_T = -\frac{\int_0^T X_s dX_s^c}{\int_0^T X_s^2 ds}. \tag{4.4}$$

Remark 4.2.1. All results in this section can also be extended to d-dimensional Ornstein-Uhlenbeck processes

$$dX_t = AX_t dt + dL_t, \quad X_0 = \tilde{X},$$

where $A \in \mathbb{R}^{d\times d}$, L is a d-dimensional Lévy process and the initial condition $\tilde{X}$ might be random or non-random. Nevertheless, we restrict our attention here to the one-dimensional case, since we are interested in the influence of the jumps of X and the role of the continuous martingale part X^c for the estimation of the drift parameter.

Now the question arises how X^c can be written in terms of the components of X under the dominating measure P^0. We consider $a_0 = 0$ such that under P^0 we have $X = L$ and $X^c = \sigma W$, where σW is the Gaussian component of L, i.e. W is a standard P^0-Wiener process. Then we know from Proposition 3.2.4 that

$$L(a_1, X^T) = \frac{dP_T^{a_1}}{dP_T^0} = \exp\left(-a_1\sigma^{-1}\int_0^T X_s\, dW_s - \frac{a_1^2}{2\sigma^2}\int_0^T X_s^2\, ds\right).$$

and it follows from Girsanov's theorem (Theorem III.3.11 in Jacod and Shiryaev [2003]) that

$$\tilde{W}_t = W_t + a_1\sigma^{-1}\int_0^t X_s\, ds$$

is a standard Wiener process under P^{a_1}. Hence, the continuous P^0-martingale part can be rewritten as

$$X_t^c = \sigma W_t = \sigma\tilde{W}_t - a_1\int_0^t X_s\, ds.$$

Thus, we can express the MLE as the true parameter a_1 minus a bias that is driven by a P^{a_1}-Wiener process:

$$\hat{a}_T = -\frac{\int_0^T X_s dX_s^c}{\int_0^T X_s^2 ds} = a_1 - \frac{\sigma\int_0^T X_s d\tilde{W}_s}{\int_0^T X_s^2 ds}.$$

The next question that comes up is can we obtain a path of the continuous part $(X_t^c(\omega))_{t\in[0,T]}$ from observations $(X_t(\omega))_{t\in[0,T]}$? By subtracting the pure jump component J of L from X, equations (3.7) and (3.8) yield that under P^{a_1}

$$\begin{aligned} X_t^c &= X_t - J_t - X_0 \\ &= X_t - \int_{\{|x|<1\}} x(N_t(dx) - t\mu(dx)) - bt - \sum_{0\le s\le t}\Delta X_s 1_{\{|\Delta X_s|\ge 1\}} - X_0, \end{aligned} \tag{4.5}$$

such that we obtain a path by path transformation from X to X^c. Hence, given a continuously observed trajectory $(X_t(\omega))_{t\in[0,T]}$ of the OU process we can use (4.5) to construct the corresponding trajectory of X^c by removing the jumps and subtracting the drift bt and the compensator $\int_{\{|x|<1\}} xt\,\mu(dx)$ of the jump measure N if this integral is finite.

The decomposition (4.5) means that under continuous observations and for general Lévy processes the drift b and the Lévy measure μ in a small neighborhood of zero are needed to evaluate the MLE, whereas σ might be unknown. When the jump part of L has finite activity the integral with respect to N in (4.5) disappears such that only the drift b has to be known in advance or could be replaced by an estimate. This situation changes when only discrete observations of X are given. Then we will use in Chapter 5 and 6 an approximation of X^c that works without knowledge of μ provided that the Blumenthal-Getoor index of L is less than two.

4.2.1 Asymptotic properties of the estimator

In this section we will show that the maximum likelihood estimator $\hat{a}_T$ for the parameter of the OU process exists uniquely and is strongly consistent, i.e. it converges to the true value almost surely as the observation length tends to infinity. We will also investigate the asymptotic distribution of $\hat{a}_T - a$ under P^a. These properties follow from the fact that the class of Lévy-driven Ornstein-Uhlenbeck processes $\{X^a : a \in \mathbb{R}_+\}$ forms a curved exponential family of stochastic processes. For a detailed discussion of the theory of exponential families of stochastic processes and their statistical applications we refer the reader to the book by Küchler and Sørensen [1997].

Let us recall the definition of LAN from Section 2.5. Let $\{P^\theta, \theta \in \Theta\}$ be a statistical experiment on a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+})$ with parameter set $\Theta \subset \mathbb{R}^d$ such that the interior $\mathring{\Theta}$ is nonempty. To prove asymptotic efficiency for the maximum likelihood estimator we will show later in this chapter that the statistical experiment corresponding to the estimation of the Ornstein-Uhlenbeck parameter is locally asymptotically normal.

Definition 4.2.2. A statistical experiment $\{P^\theta, \theta \in \Theta\}$ on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\in\mathbb{R}_+})$ is called locally asymptotically normal if for all $\theta \in \Theta$:

1. There exist $(\mathcal{F}_t)$-adapted processes $(Z_t(\theta))_{t\in\mathbb{R}_+}$ and $(T_t(\theta))_{t\in\mathbb{R}_+}$ such that for all real sequences $(\delta_t)_{t\in\mathbb{R}_+}$ with $\delta_t \to 0$ as $t \to \infty$ it holds that
$$\log\left(\frac{dP_t^{\theta+\delta_t h}}{dP_t^\theta}\right) - hT_t(\theta)^{1/2}Z_t(\theta) - \frac{1}{2}h^2T_t(\theta) = o_{P^\theta}(1)$$
for every $h \in \mathbb{R}$.

2. The convergence
$$(Z_t(\theta), T_t(\theta)) \xrightarrow{\mathcal{D}} (Z, T)$$
holds as $t \to \infty$ for a standard normal random variable Z and fixed $T \in \mathbb{R}$.

The next result states that the drift estimator is strongly consistent.

Theorem 4.2.3. *Under the condition $\sigma^2 > 0$ the maximum likelihood estimator $\hat{a}_T$ for the coefficient of the Ornstein-Uhlenbeck process based on observations from the time interval $[0,T]$ exists and is given by*
$$\hat{a}_T = -\frac{\int_0^T X_{s-}dX_s^c}{\int_0^T X_s^2 ds}, \tag{4.6}$$
where X^c denotes the continuous P^0-martingale part of X. Furthermore, under P^a the maximum likelihood estimator is unique and
$$\hat{a}_T \xrightarrow{a.s.} a$$
under P^a as $T \to \infty$.

Proof: By setting

$$
\begin{aligned}
A_t &= \frac{1}{\sigma^2}\int_0^t X_s \, dX_s^c \\
\kappa &= \frac{a^2}{2\sigma^2} \\
S_t &= \int_0^t X_s^2 \, ds
\end{aligned}
$$

we obtain a non-decreasing continuous random process $S : \Omega \times \mathbb{R}_+ \to \mathbb{R}$, a predictable continuous process $A : \Omega \times \mathbb{R}_+ \to \mathbb{R}$ and a deterministic function $\kappa : (0, \infty) \to \mathbb{R}$ on the set of parameters. With this notation the likelihood function exhibits a decomposition of the form (2.17) and we see that the class of OU processes indexed by the coefficients forms a curved exponential family.

Since $\sigma^2 > 0$, we have $S_T > 0$ P^a-a.s. for $T > 0$ and $a \in \mathbb{R}$. Hence, A_T/S_T is a.s. well defined for $T > 0$. We see immediately that κ is a steep function in the sense of Section 2.2 in Küchler and Sørensen [1997] such that the statement of the theorem follows from 5.2.1 in Küchler and Sørensen [1997]. □

To prove a central limit theorem for $\hat{a}_T$ we make the following assumptions on the drift coefficient and the Lévy measure that assure ergodicity properties of X.

Assumption 4.2.4. Suppose that $a, \sigma^2 > 0$, the Lévy measure of L satisfies

$$
\int_{|x|>1} \log|x| \; \mu(dx) < \infty \tag{4.7}
$$

and that the stationary distribution F of X has a finite second moment.

The existence of a stationary distribution under $a > 0$ and (4.7) was shown in Sato and Yamazato [1984] such that

$$
X_t \xrightarrow{\mathcal{D}} X_\infty \sim F \qquad \text{as } t \to \infty.
$$

The next result provides a central limit theorem for $\hat{a}_T$.

Theorem 4.2.5. *Suppose that Assumption 4.2.4 holds. Then under P^a*

$$
\sqrt{T}(\hat{a}_T - a) \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2}{E_a[X_\infty^2]}\right)
$$

and

$$
\sigma^{-1} S_T^{1/2}(\hat{a}_T - a) \xrightarrow{\mathcal{D}} N(0,1) \tag{4.8}
$$

as $T \to \infty$ and $S_T = \int_0^T X_s^2 \, ds$.

Remark 4.2.6. (i) When σ^2 is known or a consistent estimator is at hand we can use (4.8) to construct confidence intervals for a.

(ii) The second moment of the invariant distribution of X in terms of the Lévy-Khintchine triplet of L was derived in Lemma 2.4.4. We find that a higher jump intensity and stronger drift of L lead to a smaller asymptotic variance of the drift estimator.

Proof: Under Assumption 4.7 it follows from Theorem 2.6 in Masuda [2007] that X is ergodic. The ergodic theorem implies now that

$$\frac{1}{t}\int_0^t X_s^2\,ds \xrightarrow{a.s.} E_a[X_\infty^2] > 0 \text{ as } t \to \infty \tag{4.9}$$

under P^a. Thus, we obtain

$$S_t \xrightarrow{a.s.} \infty \quad \text{ as } t \to \infty$$

under P^a. Then, by 5.2.3 in Küchler and Sørensen [1997]

$$\sqrt{t}(\hat{a}_T - a) \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2}{E_a[X_\infty^2]}\right)$$

as $t \to \infty$. The second convergence follows immediately from (4.9) and Slutsky's lemma. □

In the next step we will investigate how the MLE performs in comparison to other estimators. In order to do this, we discuss the asymptotic behavior of the likelihood function and prove that the statistical model $\{P^a, a \in \mathbb{R}\}$ of Ornstein-Uhlenbeck processes is locally asymptotically normal. Some basic material on convergence of statistical experiments and local asymptotic theory can be found in Section 2.5.

Theorem 4.2.7. *Suppose that Assumption 4.2.4 is satisfied, then the following holds:*

1. *The statistical experiment $\{P^a, a \in \mathbb{R}_+\}$ is locally asymptotically normal.*

2. *The estimator $\hat{a}_T$ is asymptotically efficient in the sense of Hájek-Le Cam (cf. Definition 2.5.4).*

Proof: Let us denote by l the log-likelihood function

$$l_t(a) = \log\left(\frac{dP_t^a}{dP_t^0}\right) = -\frac{a}{\sigma^2}\int_0^t X_s\,dX_s^c - \frac{a^2}{2\sigma^2}\int_0^t X_s^2\,ds.$$

Then we obtain

$$\begin{aligned} l_t(a + t^{-1/2}h) - l_t(a) &= -\frac{a + t^{-1/2}h}{\sigma^2}\int_0^t X_s\,dX_s^c - \frac{(a + t^{-1/2}h)^2}{2\sigma^2}\int_0^t X_s^2\,ds \\ &\quad + \frac{a}{\sigma^2}\int_0^t X_s\,dX_s^c + \frac{a^2}{2\sigma^2}\int_0^t X_s^2\,ds \\ &= -\frac{t^{-1/2}h}{\sigma^2}\int_0^t X_s\,dX_s^c - \frac{at^{-1/2}h}{\sigma^2}\int_0^t X_s^2\,ds - \frac{t^{-1}h^2}{2\sigma^2}\int_0^t X_s^2\,ds. \end{aligned}$$

The continuous P^0-martingale part X^c can be written as

$$X^c = \sigma\tilde{W} - a\int X_s\,ds,$$

where $\tilde{W}$ is a standard P^a-Wiener process such that

$$M_t = -\frac{t^{-1/2}h}{\sigma^2}\int_0^t X_s \, dX_s^c - \frac{at^{-1/2}h}{\sigma^2}\int_0^t X_s^2 \, ds = -\frac{t^{-1/2}h}{\sigma}\int_0^t X_s \, d\tilde{W}_s$$

is a martingale under P^a. The ergodic theorem implies that

$$\frac{t^{-1}h^2}{2\sigma^2}\int_0^t X_s^2 \, ds \xrightarrow{a.s.} \frac{h^2 E_a[X_\infty^2]}{2\sigma^2}$$

for $t \to \infty$. Furthermore, Fubini's theorem yields for $t \to \infty$ that

$$t^{-1}E_a\left[\int_0^t X_s^2 \, ds\right] \to E_a[X_\infty^2].$$

It follows now from Theorem A.7.7 in Küchler and Sørensen [1997] that

$$\left(-\frac{t^{-1/2}}{\sigma^2}\int_0^t X_s \, dX_s^c - \frac{at^{-1/2}}{\sigma^2}\int_0^t X_s^2 \, ds, \frac{t^{-1}}{2\sigma^2}\int_0^t X_s^2 \, ds\right) \xrightarrow{\mathcal{D}} \left(C^{1/2}Z, C\right) \tag{4.10}$$

for $t \to \infty$ where Z is a standard normal random variable and $C = E_a[X_\infty^2]/\sigma^2$. The statement of the theorem follows now, since we have shown that

$$l_t(a + t^{-1/2}h) - l_t(a) \xrightarrow{\mathcal{D}} hC^{1/2}Z + \frac{1}{2}h^2 C$$

for $t \to \infty$. Asymptotic efficiency is a direct consequence of (4.10) and Theorem 2.5.2. □

4.2.2 The non-ergodic case

The normal convergence in Theorem 4.2.5 holds when X has ergodic properties. This is the case when $a > 0$ and

$$\int_{|x|>1} \log(x) \, \mu(dx) < \infty$$

such that a stationary solution exists (cf. Masuda [2007]). In this section we investigate the non-ergodic situation. We will only discuss the case $a < 0$ here. Assume for simplicity that L has bounded jumps and $X_0 = x \in \mathbb{R}$ is deterministic. In this setting the process $|X|$ growth exponentially and the maximum likelihood estimator (4.4) is strongly consistent by Theorem 4.2.3.

To derive the asymptotic distribution of $\hat{a}_T - a$ with a proper scaling consider the following decomposition of an exponential scaling of X:

$$\begin{aligned} H_t = e^{at}X_t - X_0 &= \int_0^t e^{as} \, dL_s = \sigma\int_0^t e^{as} \, dW_s + b\int_0^t e^{as} \, ds + \int_0^t e^{as} \, dJ_s \\ &= H_t^1 + H_t^2 + H_t^3, \end{aligned}$$

where J denotes the jump component and W the Gaussian component of L. The process H_t^1 is a zero-mean, uniformly integrable P^a-martingale such that the martingale convergence theorem yields

$$H_t^1 \xrightarrow{a.s.} H_\infty^1 \quad \text{as } t \to \infty \tag{4.11}$$

under P^a. For H^2 we obtain

$$H_t^2 = \frac{b}{a}(e^{at} - 1) \to -\frac{b}{a} = H_\infty^2 \quad \text{as } t \to \infty. \tag{4.12}$$

Denote by N the Poisson random measure corresponding to J and by $\tilde{N}(dx, dt) = N(dx, dt) - \mu(dx)dt$ its compensated version. Since L has bounded jumps, the process H^3 can be written as

$$H_t^3 = \int_0^t e^{as}\, dJ_s = \int_0^t \int_{|x|<1} e^{as} x\, \tilde{N}(dx, dt).$$

Proposition 2.2.7 yields for all $t \geq 0$ that

$$E\left[(H_t^3)^2\right] = E\left[\left(\int_0^t \int_{|x|<1} e^{as} x\, \tilde{N}(dx, dt)\right)^2\right] = \frac{e^{2at} - 1}{2a} \int_{|x|<1} x^2\, \mu(dx) < \infty.$$

Hence, H^3 is a zero-mean, uniformly integrable martingale and the martingale convergence theorem implies

$$H_t^3 \xrightarrow{a.s.} H_\infty^3 \quad \text{as } t \to \infty \tag{4.13}$$

under P^a. The convergences (4.11), (4.12) and (4.13) yield

$$H_t \xrightarrow{a.s.} H_\infty^1 + H_\infty^2 + H_\infty^3$$

under P^a and

$$e^{2at} X_t^2 \xrightarrow{p} \left(X_0 + H_\infty^1 + H_\infty^2 + H_\infty^3\right)^2.$$

The integral version of Toeplitz lemma (cf. Küchler and Sørensen [1997], Appendix B) leads to

$$e^{2at} \int_0^t X_t^2\, ds \xrightarrow{p} (-2a)^{-1} \left(X_0 + H_\infty^1 + H_\infty^2 + H_\infty^3\right)^2$$

such that for $S_t = \int_0^t X_s^2\, ds$ we can apply Theorem 5.2.2 in Küchler and Sørensen [1997] to obtain

$$S_t^{1/2}\, (\hat{a}_t - a) \xrightarrow{\mathcal{D}} N(0, 1) \quad \text{as } t \to \infty$$

under P^a. This implies under P^a that

$$e^{-at}\, (\hat{a}_t - a) \xrightarrow{\mathcal{D}} N(0, (-2a) \left(X_0 + H_\infty^1 + H_\infty^2 + H_\infty^3\right)^{-2}) \quad \text{as } t \to \infty,$$

where the limit distribution is a mixed normal distribution. The non-ergodic case leads therefore to a non-regular estimation problem with exponential rate of convergence for the maximum likelihood estimator. This is not surprising, since the exponential growth of X is due to a very strong drift such that in this setting the dynamics of the process are dominated by the drift, which simplifies its estimation considerably. A similar behavior of the likelihood estimator was found for non-ergodic Gaussian Ornstein-Uhlenbeck processes in Feigin [1976].

4.2.3 Influence of jumps

In this section we investigate how the jumps of X influence the drift estimator in the sense that if we replace the continuous martingale part X^c by $X^{cj}(\epsilon) = X^c + X^j(\epsilon)$, where

$$X_t^j(\epsilon) = \int_{|x|<\epsilon} x(N_t(dx) - t\mu(dx)), \tag{4.14}$$

for $\epsilon \in [0, \infty]$, the resulting estimate given by

$$\tilde{a}_T = -\frac{\int_0^T X_{s-} dX^{cj}(\epsilon)_s}{\int_0^T X_s^2 ds} \tag{4.15}$$

remains strongly consistent and we can show that the jumps lead to an increase in asymptotic variance that is proportional to the intensity of the jumps. For discrete observations of X, as considered in Chapter 5 and 6, the results in this section help us to understand how the jumps influences the estimation error. When we approximate X^c from discrete observations via jump filtering, it will be much simpler to detect large jumps (in Chapter 5 we will make precise what large means). In this sense $\tilde{a}_T$ mimics the situation that we are able to detect jumps larger than ϵ whereas smaller jumps cannot be removed and lead to an increased asymptotic variance as we will see in Theorem 4.2.10.

The results in this section are meant as a preparation for the discussion on discrete observations later on. We do not strive here to work under the most general conditions, but impose stationarity and slightly stronger moment conditions than necessary to simplify the presentation at some points.

Theorem 4.2.8. *Let us assume that X is stationary with $E_a[X_0^2] < \infty$ and $\sigma^2 > 0$. Then $\tilde{a}_T \to a$ with P^a-probability one as $T \to \infty$.*

Remark 4.2.9. Let us mention here that we can recover a path of $X^{cj}(\epsilon)$ from time-continuous observations of X, since X has only a finite number of jumps larger than ϵ on compact intervals and these jumps are summable such that we can simply subtract the sum of jumps larges than ϵ from X. Hence, X^{cj} is observable when time-continuous observations are given.

The assumption of a finite second moment for X implies in particular that the jump part of L exhibits a second moment (cf. Lemma 2.4.4) such that we can use in the following that

$$\int_{\mathbb{R}} x^2 \, \mu(dx) < \infty.$$

Proof: Under P^a the estimator $\tilde{a}_T$ can be written as

$$\tilde{a}_T = a + \frac{\int_0^T X_{s-} dX_s^j(\epsilon)}{\int_0^T X_s^2 ds} + \frac{\sigma \int_0^T X_s dW_s}{\int_0^T X_s^2 ds}. \tag{4.16}$$

From Theorem 4.2.3 we know that

$$\frac{\int_0^T X_s dW_s}{\int_0^T X_s^2 ds} \to 0$$

P^a-a.s. as $T \to \infty$. It remains to show that the second term in (4.16) goes to zero P^a-a.s. The quadratic variation of $X^j(\epsilon)$ is given by

$$[X^j(\epsilon)]_T = \sum_{t \in [0,T]} \Delta L_t^2 \mathbf{1}_{\{|\Delta L_t| < \epsilon\}},$$

where we sum over all jump times of L (recall that the squared jumps of a Lévy process are summable) . This yields

$$\begin{aligned} E_a \int_0^T X_{t-}^2 \, d[X^j(\epsilon)]_t &= E_a \sum_{t \in [0,T]} X_{t-}^2 \Delta L_t^2 \mathbf{1}_{\{|\Delta L_t| < \epsilon\}} \\ &= E_a \int_{[0,T] \times \{|x| < \epsilon\}} X_{t-}^2 x^2 \, N(dt, dx). \end{aligned}$$

Since $X_{t-}^2 x^2$ is square-integrable and predictable, by Lemma 4.1.4 in Applebaum [2009] we can find $0 \leq t_1 < \ldots < t_n \leq T$, $A_i \in \mathcal{B}(\mathbb{R})$ and define a predictable sequence

$$f_n(t, x) = \sum_{i=1}^n X_{t_i-}^2 x_i^2 \mathbf{1}_{(t_i, t_{i+1}]}(t) \mathbf{1}_{A_i}(x),$$

of simple functions such that f_n converges to $f(t,x) = X_{t-}^2 x^2 \mathbf{1}_{|x|<\epsilon}$ in $L^2(\Omega \times \mathbb{R}_+ \times \mathbb{R})$. Hence,

$$\begin{aligned} E_a \int_{[0,T] \times \{|x| < \epsilon\}} X_{t-}^2 x^2 \, N(dt, dx) &= \lim_{n \to \infty} E_a \int_{[0,T] \times \{|x| < \epsilon\}} f_n(t,x) \, N(dt, dx) \\ &= \lim_{n \to \infty} \sum_{i=1}^n E_a \int_{(t_i, t_{i+1}] \times A_i} X_{t_i-}^2 x_i^2 \, N(dt, dx) \\ &= \lim_{n \to \infty} \sum_{i=1}^n E_a \left[X_{t_i-}^2 x_i^2 \, N((t_i, t_{i+1}], A_i) \right]. \end{aligned}$$

Since $N((t_i, t_{i+1}], A_i)$ is independent of $\mathcal{F}_{t_i}$ and X is stationary, we obtain

$$\begin{aligned}\lim_{n\to\infty}\sum_{i=1}^{n} E_a\left[X_{t_i-}^2 x_i^2\ N((t_i,t_{i+1}],A_i)\right] &= \lim_{n\to\infty}\sum_{i=1}^{n} E_a\left[X_{t_i-}^2\right] x_i^2 E\left[N((t_i,t_{i+1}],A_i)\right]\\ &= E_a\left[X_0^2\right]\lim_{n\to\infty}\sum_{i=1}^{n} x_i^2\lambda((t_i,t_{i+1}])\mu(A_i)\\ &= E_a\left[X_0^2\right]\lim_{n\to\infty}\sum_{i=1}^{n}\int_{(t_i,t_{i+1}]\times A_i} x_i^2\lambda(dt)\mu(dx)\\ &= E_a\left[X_0^2\right] T\int_{\{|x|<\epsilon\}} x^2\ \mu(dx),\end{aligned}$$

where we used that if λ denotes the Lebesgue measure on $\mathbb{R}$, then $E[N(dt,dx)] = \mu(dt)\lambda(dx)$.

It follows now from Corollary 3 on p.73 of Protter [2004] that $M = X_- \cdot X^j(\epsilon) = \int_0^\cdot X_{t-}\ dX^j(\epsilon)_t$ is a square integrable martingale, i.e. $E[M_t^2] < \infty$ for all $t > 0$.

By Theorem I.4.40 in Jacod and Shiryaev [2003] we obtain for the predictable quadratic variation of M

$$\langle M\rangle = \langle X_- \cdot X^j(\epsilon)\rangle = X_-^2 \cdot \langle X^j(\epsilon)\rangle,$$

and by Theorem II.1.33 in Jacod and Shiryaev [2003]

$$\langle X^j(\epsilon)\rangle_t = \left\langle \int_{|x|<\epsilon} x(N_t(dx) - t\mu(dt))\right\rangle_t = t\int_{|x|<\epsilon} x^2\mu(dx).$$

Hence,

$$\langle M\rangle_T = C\int_0^T X_t^2\ dt,$$

where $C = \int_{|x|\le\epsilon} x^2\mu(dt)$. The strong law of large numbers for martingales (see Liptser [1980]) yields now

$$\frac{M_T}{\langle M\rangle_T} = \frac{\int_0^T X_{s-}dX^j(\epsilon)_s}{C\int_0^T X_s^2 ds} \stackrel{T\to\infty}{\longrightarrow} 0 \quad P^a\text{-a.s.}$$

such that finally

$$\frac{\int_0^T X_{s-}dX^{cj}(\epsilon)_s}{\int_0^T X_s^2 ds} \stackrel{T\to\infty}{\longrightarrow} a \quad P^a\text{-a.s.}$$

This completes the proof. □

The next theorem proves asymptotic normality in the case that the data is polluted by jumps. The asymptotic variance tells us the price we have to pay for taking X^{cj} instead of X^c. For the prove we need a restriction on X^j. Suppose that X^j is a spectrally negative α-stable process with $1 < \alpha \le 2$ (see Sato [1999]) with

$$E[e^{uX_1^j}] = \exp(\alpha^{-1}u^\alpha).$$

This class of X^j's is sufficiently rich to understand the influence of small jumps of L on the estimator.

Theorem 4.2.10. *Let X be a stationary Ornstein-Uhlenbeck process with $E_a[X_0^4] < \infty$, then $\tilde{a}_T$ is asymptotically normal, i.e.,*

$$\sqrt{T}(\tilde{a}_T - a) \xrightarrow{\mathcal{D}} N(0, \Sigma) \quad \textit{under } P^a \textit{ as } T \to \infty,$$

where

$$\Sigma = E_a[X_0^2]^{-1} \left(\sigma^2 + \int_{|x|<\epsilon} x^2 \, \mu(dx) \right).$$

Proof: Consider the following square integrable martingale

$$M_T = \int_0^T X_{s-} \, dX^{cj}(\epsilon)_s.$$

Since X^c is continuous, whereas $X^j(\epsilon)$ is purely discontinuous, the quadratic variation is given by

$$\begin{aligned} [M]_T &= \sigma^2 \left[\int_0^\cdot X_s \, dW_s \right]_T + \left[\int_0^\cdot X_{s-} \, dX^j(\epsilon)_s \right]_T \\ &= \sigma^2 \int_0^T X_s^2 \, ds + \sum_{t \in [0,T]} X_{t-}^2 \Delta L_t^2 1_{\{|\Delta L_t| < \epsilon\}}, \end{aligned}$$

where the sum is over all jump times of L, which is countable owing to the càdlàg property of the paths of L. We have seen in the proof of Theorems 4.2.5 and 4.2.8 that

$$\frac{1}{T} \int_0^T X_s^2 \, ds \to E_a[X_0^2] \qquad P_a\text{-a.s. as } T \to \infty,$$

and that

$$\frac{1}{T} E_a \sum_{t \in [0,T]} X_{t-}^2 \Delta L_t^2 \mathbf{1}_{\{|\Delta L_t| < \epsilon\}} = E_a[X_0^2] \int_{|x|<\epsilon} x^2 \, \mu(dx).$$

Using (4.2.3) and Theorem 4.2.3 in Applebaum [2009] the variance of the jump part equals

$$\begin{aligned}
&\frac{1}{T^2}\operatorname{Var}\left(\sum_{t\in[0,T]} X_{t-}^2 \Delta L_t^2 \mathbf{1}_{\{|\Delta L_t|<\epsilon\}}\right)\\
&= \frac{1}{T^2}\operatorname{Var}\left(\int_{[0,T]\times\{|x|<\epsilon\}} X_{t-}^2 x^2 \; N(dt,dx)\right)\\
&= \frac{1}{T^2}\operatorname{Var}\left(\int_{[0,T]\times\{|x|<\epsilon\}} X_{t-}^2 x^2 \; (N(dt,dx)-\mu(dx)\otimes\lambda(dt))\right)\\
&= \frac{1}{T^2} E\left[\left(\int_{[0,T]\times\{|x|<\epsilon\}} X_{t-}^2 x^2 \; (N(dt,dx)-\mu(dx)\otimes\lambda(dt))\right)^2\right]\\
&= \frac{1}{T^2}\int_{[0,T]\times\{|x|<\epsilon\}} E_a[X_{t-}^4 x^4]\; \mu(dx)\otimes\lambda(dt)\\
&= E_a[X_0^4]T^{-1}\int_{\{|x|<\epsilon\}} x^4\mu(dx).
\end{aligned}$$

Here again λ denotes the Lebesgue measure on $\mathbb{R}$. Hence,

$$\frac{1}{T^2}\operatorname{Var}\sum_{t\in[0,T]} X_{t-}^2 \Delta L_t^2 \mathbf{1}_{\{|\Delta L_t|<\epsilon\}} \to 0 \;\text{ as } T\to\infty.$$

This yields the stochastic convergence

$$\frac{1}{T}[M]_T \xrightarrow{p} \sigma^2 E_a[X_0^2] + E_a[X_0^2]\int_{|x|<\epsilon} x^2\; \mu(dx).$$

Since X^{cj} has bounded jumps, the maximal inequality for Ornstein-Uhlenbeck type processes in Novikov [2003] yields

$$E_a[\sup_{s\leq T}|\Delta M_s|] = E_a[\sup_{s\leq T}|\Delta\int_0^s X_{u-}\; dX^j(\epsilon)_u|] \leq \epsilon E_a[\sup_{s\leq T}|X_s|] = O(\log(T))$$

such that

$$\frac{1}{\sqrt{T}} E_a(\sup_{s\leq T}|\Delta M_s|) \to 0 \;\text{ as } T\to\infty.$$

It follows now by Theorem A.7.7 in Küchler and Sørensen [1997] that

$$\frac{M_T}{\sqrt{T}} \xrightarrow{\mathcal{D}} N(0,\Pi) \;\text{ as } T\to\infty,$$

where

$$\Pi = \sigma^2 E_a[X_0^2] + E_a[X_0^2]\int_{|x|<\epsilon} x^2\; \mu(dx).$$

Thus, we obtain by Slutsky's lemma

$$\sqrt{T}\frac{\int_0^T X_{s-}dX^{cj}(\epsilon)_s}{\int_0^T X_s^2 ds} = \frac{\frac{1}{\sqrt{T}}\int_0^T X_{s-}dX^{cj}(\epsilon)_s}{\frac{1}{T}\int_0^T X_s^2 ds} \xrightarrow{\mathcal{D}} N(0,\Sigma) \text{ as } T \to \infty,$$

where

$$\Sigma = \sigma^2 E_a[X_0^2]^{-1} + E_a[X_0^2]^{-1}\int_{|x|<\epsilon} x^2\, \mu(dx).$$

□

Comparing the asymptotic variance of $\tilde{a}_T$ to the results in Theorem 4.2.5 for $\hat{a}_T$, we see that in the presence of jumps the increase in variance depends on the intensity of jumps smaller then ϵ, which is given by

$$E_a[X_0^2]^{-1}\int_{|x|<\epsilon} x^2\, \mu(dx). \tag{4.17}$$

We will need this result later when we approximate the continuous martingale part in the case of discrete observations. This can be done by setting large increments to zero, since the increments of the jumps of X and its continuous component tend to zero at a different rate, when the step size between observations tends to zero. Hence, for a small step size with high probability only jumps smaller then ϵ remain after this transformation. Thus, when the step size tends to zero in a high frequency setting we can let ϵ go to zero such that the additional variance (4.17) vanishes asymptotically.

Formally, if the maximal jump size $\epsilon(T)$ tends to zero as $T \to \infty$, we see from Theorem 4.2.10 that $\hat{a}_T$ and $\tilde{a}_T$ have in the limit the same variance. This fact will be of importance when we investigate efficiency questions of our method.

4.2.4 Asymptotic properties of the discretized drift estimator

In this section we discuss the limiting behavior of the discretized drift estimator. Our objective is here to investigate the influence of the discretization error on the estimation error. We will give asymptotic conditions on the maximal observation distance in order to obtain an efficient estimator. Since we assume throughout that increments of the continuous martingale part are given, we obtain a pseudo estimator that cannot be applied in practice without further ado. Nevertheless, besides some theoretical interest we will need the convergence results form this section in Chapter 5 when we approximate the continuous martingale part and construct an efficient estimator based on discrete observations.

High-frequency asymptotics

First, we will discuss high frequency asymptotics for arbitrarily spaced observations

$$X_{t_0}, \ldots, X_{t_n} \text{ where } 0 = t_0 \le \ldots \le t_n = T$$

such that

$$\lim_{n\to\infty} \max_{0\le i\le n-1}\{t_{i+1} - t_i | 1 \le i \le n\} = 0.$$

Let $\Delta_n = \max\{t_{i+1} - t_i | 0 \le i \le n-1\}$ and discretize the MLE as follows

$$\breve{a}_n = -\frac{\sum_{i=0}^{n-1} X_{t_i} \Delta_i X^c}{\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i}$$

where $\Delta_i X^c = X_{i+1}^c - X_i^c$ and $\Delta_i = t_{i+1} - t_i$. Our first result tells us that, if Δ_n tends to zero sufficiently fast, the discretized MLE and estimator $\hat{a}_T$ based on continuous observations have the same asymptotic distribution.

Proposition 4.2.11. *Assume that X is stationary and $E(X_0^2) < \infty$. If $T_n \Delta_n = o(1)$ then*

$$\sqrt{T_n}(\breve{a}_n - a) \xrightarrow{\mathcal{D}} N(0, \sigma^2 E_a[X_0^2]^{-1})$$

under P^a. Hence, under these conditions the discretized MLE $\breve{a}_n$ and the MLE $\hat{a}_T$ based on continuous observations converge to the same asymptotic distribution as $T \to \infty$.

Proof: Let W denotes a P^a-Wiener process. The continuous P^0-martingale part can be written as

$$X_t^c = \sigma W_t - a \int_0^t X_s \, ds.$$

This leads to the decomposition

$$T_n^{1/2}(\breve{a}_n - a) = T_n^{1/2} a \left(\frac{\sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} X_s \, ds}{\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i} - 1 \right) - T_n^{1/2} \frac{\sigma \sum_{i=0}^{n-1} X_{t_i} \Delta_i W}{\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i} = S_n^1 - S_n^2.$$

We will show now that $S_n^1 \xrightarrow{p} 0$ and $S_n^2 \xrightarrow{\mathcal{D}} N(0, \sigma^2 E_a[X_0^2]^{-1})$ as $n \to \infty$ such that the statement of the proposition follows. Define $\lfloor t \rfloor_n = max_{i \le n}\{t_i | t_i \le t\}$. Let us first consider convergence of S_n^1. Observe that

$$S_n^1/a = \frac{T_n^{-1/2}(\sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} X_s \, ds - \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i)}{T_n^{-1} \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i}. \tag{4.18}$$

For the numerator we obtain

$$\begin{aligned} T_n^{-1/2} E\left[\left| \sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} X_s \, ds - \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \right| \right] &\le T_n^{-1/2} \int_0^{T_n} E_a \left[|X_{\lfloor t \rfloor_n} X_t - X_{\lfloor t \rfloor_n}^2| \right] \, dt \\ &= O(T_n^{1/2} \Delta_n^{1/2}) \end{aligned} \tag{4.19}$$

such that the numerator converges to zero in L^1. A similar estimate for the denominator yields

$$T_n^{-1} E_a \left[\left| \int_0^{T_n} X_t^2 \, dt - \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \right| \right] = O(\Delta_n^{1/2}),$$

and since the ergodic theorem implies that $T_n^{-1}\int_0^{T_n} X_t^2\, dt \xrightarrow{p} E_a[X_0^2]$ as $n \to \infty$, we conclude

$$T_n^{-1}\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \xrightarrow{p} E_a[X_0^2] \tag{4.20}$$

as $n \to \infty$. This convergence together with (4.18) and the estimate (4.19) imply that $S_n^1 \xrightarrow{p} 0$ as $n \to \infty$.

It remains to prove convergence of S_n^2. From Itô's isometry and stationarity of X we obtain for the numerator of S_n^1 that

$$\begin{aligned}
T_n^{-1} E_a &\left[\left(\int_0^{T_n} X_t\, dW_t - \sum_{i=0}^{n-1} X_{t_i}\Delta_i W\right)^2\right] = T_n^{-1} E_a\left[\left(\int_0^{T_n} (X_t - X_{\lfloor t\rfloor_n})\, dW_t\right)^2\right] \\
&= T_n^{-1} E_a\left[\int_0^{T_n} (X_t - X_{\lfloor t\rfloor_n})^2\, dt\right] = T_n^{-1}\int_0^{T_n} E_a\left[(X_t - X_{\lfloor t\rfloor_n})^2\right]\, dt \\
&\leq 2T_n E_a[X_0^2 - X_0 X_{\Delta_n}] = 2T_n E_a[X_0^2](1 - e^{-a\Delta_n}) = O(\Delta_n).
\end{aligned}$$

The numerator of S_n^2 is a continuous martingale and its quadratic variation converges due to the ergodic theorem to the second moment of X. The martingale central limit theorem implies now

$$T_n^{-1/2}\sigma \int_0^{T_n} X_t\, dW_t \xrightarrow{\mathcal{D}} N(0, \sigma^2 E_a(X_0^2))$$

such that also

$$T_n^{-1/2}\sigma \sum_{i=0}^{n-1} X_{t_i}\Delta_i W \xrightarrow{\mathcal{D}} N(0, \sigma^2 E_a(X_0^2))$$

as $n \to \infty$. This convergence together with (4.20) and Slutsky's lemma lead to

$$S_n^1 \xrightarrow{\mathcal{D}} N(0, \sigma^2 E_a[X_0^2]^{-1})$$

as $n \to \infty$. This completes the proof. □

Long time asymptotics

Now we will turn our attention to the long time asymptotics of $\hat{a}_n$ on an equidistant grid $\Delta, 2\Delta, \ldots, n\Delta$ for $\Delta \in \mathbb{R}_+$ fixed and let $n \to \infty$. In this situation the estimator exhibits a bias that tends to zero as Δ becomes small and by using a bias correction we can reduce the order of the asymptotic bias to $O(\Delta^2)$. Our results are closely related to the work by Küchler and Sørensen [2010] on parameter estimation for affine stochastic delay differential equations.

Given discrete observations $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ a discretized version of $\hat{a}$ is given by

$$\hat{a}_n = -\frac{A_n}{I_n},$$

where

$$A_n = \sum_{m=0}^{n-1} X_{m\Delta}\, \delta X_m^c$$

and

$$I_n = \Delta \sum_{m=0}^{n-1} X_{m\Delta}^2$$

with increments $\delta X_m^c = X_{(m+1)\Delta}^c - X_{m\Delta}^c$. Let a denote the true parameter of equation (2.13).

Theorem 4.2.12. *On the assumption that X is stationary with $E_a(X_0^2) < \infty$ the drift estimator satisfies*

$$\hat{a}_n + \frac{\hat{a}_n^2}{2}\Delta \overset{n\to\infty}{\longrightarrow} a + O(\Delta^2) \qquad P_a\text{-}a.s.$$

Proof: We have seen in Section 4.2 that when W is a P^a-Wiener process we can write the continuous P^0-martingale part as

$$X_t^c = \sigma W_t - a \int_0^t X_s\, ds.$$

Therefore, we can rewrite A_n as

$$\begin{aligned} A_n &= \sigma \sum_{m=0}^{n-1} X_{m\Delta}\, \delta W_m - a \sum_{m=0}^{n-1} X_{m\Delta} \int_{m\Delta}^{(m+1)\Delta} X_s\, ds \\ &= \sigma \sum_{m=0}^{n-1} X_{m\Delta}\, \delta W_m - a \sum_{m=0}^{n-1} X_{m\Delta} \int_{m\Delta}^{(m+1)\Delta} (X_s - X_{m\Delta})\, ds - a\Delta \sum_{m=0}^{n-1} X_{m\Delta}^2 \\ &=: Z_n - aR_n - aI_n \end{aligned} \tag{4.21}$$

From Theorem 4.3 in Masuda [2004] we know that X is β-mixing and hence geometrically ergodic. Therefore, the ergodic theorem implies the P^a-almost sure convergence of

$$\frac{I_n}{n} \overset{n\to\infty}{\longrightarrow} \Delta E_a(X_0^2), \qquad \frac{Z_n}{n} \overset{n\to\infty}{\longrightarrow} 0,$$

and

$$\frac{R_n}{n} \overset{n\to\infty}{\longrightarrow} E_a\left[\int_{m\Delta}^{(m+1)\Delta} \left(X_{m\Delta}X_t - X_{m\Delta}^2\right)\, dt\right] = \int_0^\Delta K_a(t)\, dt - \Delta E_a[X_0^2],$$

where $K_a(t) = E_a[X_0 X_t] = E_a[X_0^2]e^{-at}$ denotes the covariance function of X. Hence, we obtain

$$\hat{a}_n = -\frac{Z_n}{I_n} + a\frac{R_n}{I_n} + a \overset{n\to\infty}{\longrightarrow} a\frac{\int_0^\Delta K_a(t)\, dt}{\Delta E_a[X_0^2]} \qquad P_a\text{-a.s.}$$

Finally,

$$a\frac{\int_0^\Delta K_a(t)\, dt}{\Delta E_a[X_0^2]} = \frac{a}{\Delta}\int_0^\Delta e^{-at}\, dt = a - \frac{a^2}{2}\Delta + O(\Delta^2)$$

□

As expected the asymptotic bias of the discretized >estimator tends to zero as Δ becomes small. Our next result shows that $\hat{a}_n$ is asymptotically normal if X is stationary and exhibits a moment of order $2+\delta$ for some $\delta > 0$. The proof is based on classical central limit theorems for mixing sequences as discussed in Doukhan [1994].

Theorem 4.2.13. *Let us assume that X is stationary, $E_a[|X_0|^{2+\delta}] < \infty$ for some $\delta > 0$. We set*

$$\mathcal{R}(a) = \int_0^\Delta K_a(t)\,dt - \Delta E_a[X_0^2],$$

where $K_a = E_a[X_0^2]e^{-at}$ denotes the covariance function of X. Then the distribution of

$$\sqrt{n}\left(\hat{a}_n - a - a\frac{\mathcal{R}(a)}{\Delta E_a[X_0^2]}\right)$$

tends as $n \to \infty$ to a centered normal distribution.

Proof: By using (4.21) and $\mathcal{R}(a) = \int_0^\Delta K_a(t)\,dt - \Delta E_a[X_0^2]$ we obtain

$$\begin{aligned}
\sqrt{n}\left(\hat{a}_n - a - a\frac{\mathcal{R}(a)}{\Delta E_a[X_0^2]}\right) &= \sqrt{n}\left(\frac{Z_n}{I_n} + a\left(\frac{R_n}{I_n} - \frac{\mathcal{R}(a)}{\Delta E_a[X_0^2]}\right)\right) \\
&= \frac{na}{I_n}\left[\frac{R_n}{\sqrt{n}} - \sqrt{n}\mathcal{R}(a) + \sqrt{n}\mathcal{R}(a)\right] + \sqrt{n}a\frac{\mathcal{R}(a)}{\Delta E_a[X_0^2]} + \frac{n}{I_n}\frac{Z_n}{\sqrt{n}} \\
&= \frac{na}{I_n}\frac{R_n - n\mathcal{R}(a)}{\sqrt{n}} - \mathcal{R}(a)\frac{na}{I_n}\frac{I_n - n\Delta E_a[X_0^2]}{\sqrt{n}\Delta E_a[X_0^2]} + \frac{n}{I_n}\frac{Z_n}{\sqrt{n}}.
\end{aligned}$$

The last expression has the same asymptotic distribution as

$$\frac{1}{\Delta E_a[X_0^2]}\left(a\frac{R_n - n\mathcal{R}(a)}{\sqrt{n}} - \mathcal{R}(a)a\Delta E_a[X_0^2]\frac{I_n - n\Delta E_a[X_0^2]}{\sqrt{n}} + \frac{Z_n}{\sqrt{n}}\right)$$

Masuda [2004] proved in Theorem 4.3 that X is exponentially β-mixing and hence also α mixing. Therefore, we see that $R_n - n\mathcal{R}(a)$, $I_n - n\Delta E_a[X_0^2]$ and Z_n are sums of centered exponentially α-mixing sequences. In order to apply the central limit theorem for α-mixing sequences (see for example Doukhan [1994], Section 1.5, Theorem 1) it remains to check that all sequences exhibit finite moments of order $2+\delta$ for some $\delta > 0$.

By equation (2.14) we obtain for the summands of Z_n

$$E_a[|X_{m\Delta}W_{m\Delta}|]^{2+\delta} = E_a\left[|e^{-am\Delta}X_0W_{m\Delta} + W_{m\Delta}\int_0^{m\Delta} e^{-a(m\Delta-s)}\,dL_s|^{2+\delta}\right].$$

The Lévy -Itô decomposition $L = \sigma W + J$ with quadratic pure jump part J and Gaussian component W yields now

$$E_a\left[|X_{m\Delta}W_{m\Delta}|^{2+\delta}\right] = E_a\left[|W_{m\Delta}|^{2+\delta}\right] E_a\left[\left|e^{-am\Delta}X_0 + \int_0^{m\Delta} e^{-a(m\Delta-s)}\,dW_s\right.\right.$$
$$\left.\left. + \int_0^{m\Delta} e^{-a(m\Delta-s)}\,dJ_s\right|^{2+\delta}\right]$$
$$\leq 3^{\frac{\delta}{2}} C E_a[|W_{m\Delta}|^{2+\delta}] E_a\left[|e^{-am\Delta}X_0|^{2+\delta}\right] + E_a\left[\left|\int_0^{m\Delta} e^{-a(m\Delta-s)}\,dW_s\right|^{2+\delta}\right]$$
$$+ E_a\left[\left|\int_0^{m\Delta} e^{-a(m\Delta-s)}\,dJ_s\right|^{2+\delta}\right] < \infty,$$

where we have used independence of X_0, W, and J and an inequality by Marcinkiewicz and Zygmund that can be found in Loeve [1977], p. 276.

For R_n we find

$$E_a\left[|R_n|^{2+\delta}\right] = E_a\left[\left|\sum_{m=0}^{n-1}\int_{m\Delta}^{(m+1)\Delta} X_{m\Delta}X_s\,ds + \Delta\sum_{m=0}^{n-1} X_{m\Delta}^2\right|^{2+\delta}\right]$$
$$\leq (2+n)^{1+\delta}\left[\sum_{m=0}^{n-1} E_a\left[\left|\int_{m\Delta}^{(m+1)\Delta} X_{m\Delta}X_s\,ds\right|^{2+\delta}\right] + \Delta\sum_{m=0}^{n-1} E_a\left[\left|X_{m\Delta}^2\right|^{2+\delta}\right]\right] < \infty,$$

where we have used a moment inequality from DasGupta [2008], p. 650. We have by assumption $E_a|Z_n|^{2+\delta} < \infty$. Hence, the asymptotic normality of all three terms

$$\frac{R_n - n\mathcal{R}(a)}{\sqrt{n}}, \quad \frac{I_n - n\Delta E_a[X_0^2]}{\sqrt{n}} \text{ and } \frac{Z_n}{\sqrt{n}}$$

follows from the central limit theorem for α-mixing sequences (Doukhan [1994], Section 1.5, Theorem 1). □

It remains to determine the variance of the limiting distribution in the central limit theorem.

Remark 4.2.14. The case of equidistant observations $X_0, X_\Delta, \ldots, X_{n\Delta}$ can also be treated by classical time series methods, since we can rewrite the observations as an AR(1) process via

$$X_{m\Delta} = \theta X_{(m-1)\Delta} + \epsilon_m; \quad m = 1, \ldots, n \tag{4.22}$$

where

$$\theta = e^{-a\Delta}$$

and

$$\epsilon_m = \int_{(m-1)\Delta}^{m\Delta} e^{-a(m\Delta-s)}\,dL_s.$$

Then, the sequence $(\epsilon_m)_{1\leq m\leq n}$ is iid due to the stationary and independent increments of L.

This approach was studied in Brockwell et al. [2007] in the case that L is a subordinator with finite second moments and on the assumption that the distribution of the ϵ_m is regularly varying at zero.

Also for general Lévy processes the AR(1) process (4.22) can be estimated by classical least squares or maximum likelihood techniques that lead to strongly consistent and asymptotically normal estimators provided that X is stationary and has finite second moments. This time series approach breaks down if the observations of X are not equidistant. In this case our approach yields equally good results as in the equidistant case provided that the maximal time step between observations is sufficiently small.

4.3 Lévy-driven square-root processes

As a second example of a jump diffusion that is widely used in applications we consider the drift estimation for square-root processes from Section 3.3. Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P)$ be a filtered probability space on which all processes will be defined in the following. Then X is a strong solution of

$$dX_t = -aX_t dt + \sigma\sqrt{X_t} dW_t + dL_t, \quad 0 \leq t < \infty,$$

driven by a Lévy process $L_t = bt + J_t$ and a standard Wiener process W, where $\sigma, b > 0$ and J is assumed to be subordinator (see Sato [1999], Section 4.21 on subordinators). The starting value is $X_0 = x > 0$ and the drift parameter $a > 0$. Under these conditions the process X stays nonnegative at all times such that the square-root is well defined (cf. Section 3.3 and the references given there). This class of processes has been used by Cox et al. [1985] for short rate dynamics in term structure modeling when L is a Brownian motion. Therefore, they are also known as Cox-Ingersoll-Ross processes in financial applications.

As a mapping from Ω to $D[0,\infty)$ the process X induces for every a a measure P^a on $D[0,\infty)$. Proposition 3.3.1 yields absolute continuity of these measures and enables us to derive an explicit maximum likelihood estimator for a by solving

$$\frac{\partial}{\partial a} \log \frac{dP_T^a}{dP_T^0} = -\frac{1}{\sigma^2} X_T^c - \frac{a}{\sigma^2} \int_0^T X_t \, dt \overset{!}{=} 0,$$

where X^c denotes the continuous P^0-martingale part. Thus, we obtain for continuously observed X the MLE

$$\hat{a}_T = -\frac{X_T^c}{\int_0^T X_t \, dt}. \tag{4.23}$$

The form of the likelihood function implies that the class of square-root processes forms a curved exponential family such that we can apply general results for exponential families to derive the asymptotic behavior of $\hat{a}_T$. Denote by (b, σ^2, μ) the Lévy -Khintchine triplet of L.

Theorem 4.3.1. *(i) Suppose $\sigma^2 > 0$, then the drift estimator $\hat{a}_T$ is strongly consistent.*

(ii) If additionally X is ergodic and has an invariant distribution such that $X_t \xrightarrow{\mathcal{D}} X_\infty$ as

$t \to \infty$ and $E_a[X_\infty] < \infty$, then

$$\sqrt{T}(\hat{a}_T - a) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2 E_a[X_\infty]^{-1}\right)$$

and

$$\sigma^{-1} S_T^{1/2}(\hat{a}_T - a) \xrightarrow{\mathcal{D}} N(0,1)$$

under P^a as $T \to \infty$ and $S_T = \int_0^T X_s \, ds$.

When a consistent estimator $\hat{\sigma}^2$ for σ is at hand, the second convergence in (ii) can be used to obtain confidence statements for a.

Proof: Let W denote the Gaussian component of L from the Lévy-Itô decomposition that is W is a P^0-Wiener process. It follows from Girsanov's theorem that the continuous P^0-martingale part can be written as

$$X_t^c = \sigma \int_0^t X_s^{1/2} \, dW_s = -a \int_0^t X_s \, ds + \sigma \int_0^t |X_s|^{1/2} \, d\tilde{W}_s,$$

where $\tilde{W}$ is a standard P^a-Wiener process. From this representation of X^c the MLE can be rewritten as

$$\hat{a}_T = -\frac{\sigma \int_0^T |X_s|^{1/2} \, d\tilde{W}_s - a \int_0^T X_s \, ds}{\int_0^T X_s \, ds} = a - \frac{\sigma \int_0^T |X_s|^{1/2} \, d\tilde{W}_s}{\int_0^T X_s \, ds}.$$

Since $\sigma^2 > 0$ and $X_s \geq 0$ for almost all $s > 0$, it follows that

$$\int_0^T X_s \, ds \to \infty \qquad \text{as } T \to \infty$$

and P^a-a.s. such that by Lemma 17.4 in Liptser and Shiryaev [2001] we obtain

$$\frac{\sigma \int_0^T |X_s|^{1/2} \, d\tilde{W}_s}{\int_0^T X_s \, ds} \to 0$$

P^a-a.s. as $T \to \infty$. This proves strong consistency of $\hat{a}_T$. It remains to prove the central limit theorem. The ergodic theorem yields

$$T^{-1} \int_0^T X_t \, dt \xrightarrow{a.s.} E_a[X_\infty] > 0 \tag{4.24}$$

under P^a as $T \to \infty$. Now we can apply Theorem 5.2.3 in Küchler and Sørensen [1997] to obtain the result

$$\sqrt{T}(\hat{a}_T - a) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2 E_a[X_\infty]^{-1}\right) \tag{4.25}$$

under P^a as $T \to \infty$. □

4.3.1 Local asymptotic normality

To show that the MLE exhibits a minimal asymptotic variance we are going to prove local asymptotic normality for the underlying statistical experiment. As a corollary we obtain that the estimator is asymptotically efficient in the sense of Le Cam theory.

Theorem 4.3.2. *Suppose that the assumptions in Theorem 4.3.1 hold. Then the following holds:*

(i) The statistical experiment $\{P^a, a \in \mathbb{R}_+\}$ *is locally asymptotically normal.*

(ii) The drift estimator $\hat{a}_T$ *is asymptotically efficient in the sense of Hájek-Le Cam.*

Proof: We have to show that the log-likelihood

$$l_t(a) = log\left(\frac{dP_t^a}{dP_t^0}\right) = -\frac{a}{\sigma^2}X_T^c - \frac{a^2}{2\sigma^2}\int_0^T X_t\, dt$$

converges locally to the likelihood of a Gaussian shift experiment. For the local parameter $a + t^{-1/2}h$ we obtain by applying Girsanov's theorem that

$$\begin{aligned} l_t(a+t^{-1/2}h) - l_t(a) &= -\frac{a+t^{-1/2}h}{\sigma^2}X_T^c - \frac{(a+t^{-1/2}h)^2}{2\sigma^2}\int_0^t X_s\, ds \\ &+ \frac{a}{\sigma^2}X_T^c + \frac{a^2}{2\sigma^2}\int_0^t X_s\, ds \\ &= -\frac{t^{-1/2}h}{\sigma^2}X_T^c - \frac{at^{-1/2}h}{\sigma^2}\int_0^t X_s\, ds - \frac{t^{-1}h^2}{2\sigma^2}\int_0^t X_s\, ds \\ &= -\frac{t^{-1/2}h}{\sigma}\int_0^t X_s^{1/2}\, d\tilde{W}_s - \frac{t^{-1}h^2}{2\sigma^2}\int_0^t X_s\, ds, \end{aligned}$$

where $\tilde{W}$ denotes a P^a-Wiener process. Next, we observe that the ergodic theorem leads to

$$t^{-1}\int_0^t X_s\, ds \xrightarrow{a.s.} \sigma^2 E_a[X_\infty]$$

for $t \to \infty$ and that Itô's isometry implies

$$\sigma^2 t^{-1} E_a\left[\left(\int_0^t X_s^{1/2}\, d\tilde{W}_s\right)^2\right] \to \sigma^2 E_a[X_\infty]$$

such that Theorem A.7.7 in Küchler and Sørensen [1997] yields

$$\left(-\frac{t^{-1/2}}{\sigma^2}\int_0^t X_s^{1/2}\, d\tilde{W}_s, \frac{t^{-1}}{2\sigma^2}\int_0^t X_s\, ds\right) \xrightarrow{\mathcal{D}} \left(C^{1/2}Z, C\right)$$

under P^a as $t \to \infty$, where Z is a standard normal random variable and $C = \sigma^2 E_a[X_0]$. From this joint convergence we finally conclude that under P^a we obtain

$$l_t(a + t^{-1/2}h) - l_t(a) \xrightarrow{\mathcal{D}} hE_a[X_\infty]^{1/2}\sigma^{-1}Z + \frac{1}{2}h^2 E_a[X_\infty]\sigma^{-2}.$$

□

4.4 Jump diffusion models with affine drift parameter

Suppose that the drift coefficient in (4.1) depends linearly on θ, i.e.,

$$\delta(\theta, s, x) = g(s, x) + \theta f(s, x)$$

for known functions $f, g : \mathbb{R}_+ \times \mathbb{R}^d \to \mathbb{R}^d$ and let $\Theta \subset \mathbb{R}^n$ be an open subset. Then $\{P^\theta, \theta \in \Theta\}$ is a curved exponential family as defined in Section 2.6 and $\mathcal{L}$ is a quadratic polynomial in θ such that (4.2) has the unique solution

$$\hat{\theta}_T = \frac{\int_0^T c(s, X_s)^{-1} f(s, X_s)\, dX_s^c}{\int_0^T f(s, X_s)^\top c(s, X_s)^{-1} f(s, X_s)\, ds},$$

where X^c denotes the continuous P^0-martingale part and we assume in the following that $c(s, x) = \gamma\Sigma\Sigma^\top\gamma^\top(s, x)$ from (4.1) is invertible for all s and x. This class of models generalizes the Ornstein-Uhlenbeck type and square-root process that we have investigated in the previous sections. Consistency and asymptotic normality for $\hat{\theta}_T$ follow under ergodicity of X from Theorem 5.2.1 and 5.2.2 in Küchler and Sørensen [1997]. We summarize the asymptotic properties of $\hat{\theta}_T$ in the following theorems.

Theorem 4.4.1. *The maximum likelihood estimator $\hat{\theta}_T$ is strongly consistent, i.e.,*

$$\hat{\theta}_T \xrightarrow{a.s.} \theta \quad \text{under } P^\theta \text{ as } t \to \infty.$$

To obtain asymptotic normality of $\hat{\theta}_T$ we impose ergodicity of X. Therefore, it is necessary that the coefficients do not explicitly depend on t, i.e. $f, g : \mathbb{R}^d \to \mathbb{R}^d$ and $\gamma : \mathbb{R}^d \to \mathbb{R}^{d\times m}$. Conditions for ergodicity of jump diffusions of this type in terms of there characteristics were developed in Masuda [2007].

Theorem 4.4.2. *Suppose that X is ergodic and has an invariant distribution such that $X_t \xrightarrow{\mathcal{D}} X_\infty$ as $t \to \infty$ and*

$$\Psi = E_\theta[f(X_\infty)^\top c(X_\infty)^{-1} f(X_\infty)] < \infty.$$

Then under P^θ we obtain

$$T^{1/2}\left(\hat{\theta}_T - \theta\right) \xrightarrow{\mathcal{D}} N(0, \Psi^{-1})$$

as $T \to \infty$.

4.5 Stochastic delay differential equations

All jump diffusion processes considered so far have the Markov property in common. In this section we will see that the likelihood approach generalizes also to a non-Markovian setting.

Let L be a Lévy process with Lévy-Khintchine triplet (γ, σ^2, μ) and consider the following linear stochastic delay differential equation (SDDE) driven by L.

$$\begin{aligned} dX_t &= aX_t\, dt + bX_{t-1}\, dt + dL_t, && t > 0, \\ X_t &= X_t^0, && t \in [-1, 0], \end{aligned} \tag{4.26}$$

where $a, b \in \mathbb{R}$ and $X^0 : [-1, 0] \times \Omega \to \mathbb{R}$ is the initial process with càdlàg trajectories that is assumed to be independent of L. For general results on existence and uniqueness of stationary solutions of delay equations we refer to Gushchin and Küchler [2000] and Mohammed and Scheutzow [1990]. A short summary of the most important results can be found in Section 2.4.

When x_0 denotes the fundamental solution of the corresponding deterministic delay equation

$$\begin{aligned} \dot{x}(t) &= ax(t)\, dt + bx(t-1)\, dt, && t > 0, \\ x(t) &= 0, && t \in [-1, 0), \\ x(0) &= 1, \end{aligned}$$

a unique solution of (4.26) is given by

$$X_t = x_0(t)X_0^0 + b\int_{-1}^{0} x_0(t-s-1)X_s^0\, ds + \int_0^t x_0(t-s)\, dL_s.$$

It was shown in Gushchin and Küchler [2000] that a stationary solution of (4.26) exists if and only if

$$\int_{|x|>1} \log|x|\, \mu(dx) < \infty \tag{4.27}$$

and when the characteristic equation $\lambda - a - be^{\lambda} = 0$ admits only solutions with negative real part, i.e.,

$$v_0 = \sup_{\lambda\in\mathbb{C}}\{\mathrm{Re}(\lambda) | \lambda - a - be^{\lambda} = 0\} < 0. \tag{4.28}$$

4.5.1 Parameter estimation

In the following we discuss the problem of estimating the parameter $\theta = (a, b) \in \mathbb{R}^2$ from continuous observations $(X_t)_{t\in[0,T]}$ for $T \in \mathbb{R}_+$. Let P^θ denote the measure induced by X on $D[-1, \infty)$. The process X is by Theorem V.3.7 in Protter [2004] a semimartingale with characteristics

$$\begin{aligned} B_t &= \gamma t + a \int_0^t X_s \, ds + b \int_0^t X_{s-1} \, ds, \\ C_t &= \sigma^2 t, \\ \nu(dt, dx) &= \lambda(dt)\mu(dx). \end{aligned}$$

Hence, the likelihood function for the experiment $(P^\theta, \theta \in \mathbb{R}^2)$ follows from (2.10) in Küchler and Sørensen [1989]:

$$L(\theta, X, T) = \frac{dP_T^\theta}{dP_T^{(0,0)}} = \exp\left(\theta^\top V_T - \frac{1}{2}\theta^\top I_T \theta\right),$$

where

$$V_T = \begin{pmatrix} \int_0^T X_t \, dX_t^c \\ \int_0^T X_{t-1} \, dX_t^c \end{pmatrix},$$

the process X^c denotes the continuous P^0-martingale part and I_T is the observed Fisher information given by

$$I_T = \begin{pmatrix} \int_0^T X_t^2 \, dt & \int_0^T X_t X_{t-1} \, dt \\ \int_0^T X_t X_{t-1} \, dt & \int_0^T X_{t-1}^2 \, dt \end{pmatrix}.$$

Hence, we are working on a set of measures $(P^\theta, \theta \in \mathbb{R}^2)$ that forms a curved exponential family in the sense of Küchler and Sørensen [1997]. If we assume that I is non-singular the maximum likelihood estimator takes the explicit form

$$\hat{\theta}_T = I_T^{-1} V_T. \tag{4.29}$$

The continuous P^0-martingale part of X is $X^c = \sigma \tilde{W}$, where $\tilde{W}$ is the Gaussian part of L and hence a P^0-Wiener process. Under the measure P^θ Girsanov's theorem implies that

$$W = \tilde{W} - \frac{a}{\sigma} \int X_t \, dt - \frac{b}{\sigma} \int X_{t-1} \, dt$$

is a P^θ-Wiener process. Hence, the continuous martingale part can be rewritten as

$$X_t^c = a \int_0^t X_t \, dt + b \int_0^t X_{t-1} \, dt + \sigma W_t. \tag{4.30}$$

This representation of X^c will be important in the following when we investigate the asymptotic behavior of the drift estimator, since we are interested in convergence results under the true measure P^θ.

4.5.2 Asymptotic properties

In this section we will prove consistency and asymptotic normality for the maximum likelihood estimator for θ.

Assumption 4.5.1. Suppose that $\sigma^2 > 0$, that (4.27) and (4.28) hold and that X is ergodic and exhibits finite second moments.

Remark 4.5.2. That $\sigma^2 > 0$ yields existence of the likelihood function and hence the existence of $\hat{\theta}_T$. The conditions (4.27) and (4.28) are necessary and sufficient for the existence of a stationary solution of (4.26) and they will be needed together with the ergodicity of X to derive the limit theorems in this section. That we are in the stationary case together with the existence of second moments of X is a requirement in order to obtain asymptotic normality for the estimation error. It is an interesting open question to derive convergence results for $\hat{\theta}_T$ in the non-stationary case. Gushchin and Küchler [1999] have investigated the asymptotic behavior of the likelihood estimator in the case of Gaussian delay equations and they found that the parameter space decomposes into eleven different regions that lead to eleven different limiting behaviors of the likelihood function. We expect a similarly complex situation for Lévy-driven delay equations.

Theorem 4.5.3. *Let Assumption 4.5.1 be satisfied, then the MLE $\hat{\theta}_T$ is strongly consistent, i.e., under P^θ*

$$\hat{\theta}_T \xrightarrow{a.s.} \theta \text{ as } t \to \infty.$$

Proof: We rewrite the MLE as follows:

$$\hat{\theta}_T = I_T^{-1} V_T = \theta + I_T^{-1} \tilde{V}_T, \tag{4.31}$$

where

$$\tilde{V}_T = \begin{pmatrix} \sigma \int_0^T X_t \, dW_t \\ \sigma \int_0^T X_{t-1} \, dW_t \end{pmatrix}$$

and W is a P^θ-Wiener process. It remains to prove that $I_T^{-1} \tilde{V}_T \xrightarrow{a.s.} 0$ as $T \to \infty$. First of all, the ergodic theorem implies that

$$T^{-1} I_T \xrightarrow{a.s.} \Sigma \quad \text{as } T \to \infty \tag{4.32}$$

under P^θ, where

$$\Sigma_{00} = \Sigma_{11} = \lim_{s \to \infty} E_\theta \left[X_s^2 \right]$$

and

$$\Sigma_{01} = \Sigma_{10} = \lim_{s \to \infty} E_\theta \left[X_s X_{s-1} \right].$$

That these limits exist follows under (4.27) and (4.28) from Theorem 3.1 in Gushchin and Küchler [2000]. Observe that the predictable quadratic variation of $\tilde{V}_T$ converges to

$$T^{-1} \sigma^2 \left\langle \int_0^\cdot X_s \, dW_s \right\rangle_T = T^{-1} \sigma^2 \int_0^T X_s^2 \, ds \xrightarrow{a.s.} \sigma^2 \Sigma_{00}$$

and

$$T^{-1}\sigma^2 \left\langle \int_0^{\cdot} X_{s-1}\, dW_s \right\rangle_T \xrightarrow{a.s.} \sigma^2 \Sigma_{00}.$$

under P^θ as $T \to \infty$. Hence, the strong law of large numbers for martingales 2.3.7 yields

$$T^{-1} \int_0^T X_s\, dW_s \xrightarrow{a.s.} 0$$

and

$$T^{-1} \int_0^T X_{s-1}\, dW_s \xrightarrow{a.s.} 0$$

under P^θ as $T \to \infty$ such that the claim follows from (4.31) and (4.32). $\square$

In the ergodic case the following central limit theorem holds for estimation error.

Theorem 4.5.4. *Suppose that Assumption 4.5.1 holds and that Σ from* (4.32) *is invertible, then*

$$T^{1/2}\left(\hat{\theta}_T - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2 \Sigma^{-1}\right).$$

Proof: Since Σ is symmetric, there exists an orthogonal matrix U such that $U^\top \Sigma U$ is a diagonal matrix. Define $Y_t = (X_t, X_{t-1})^\top$ and

$$Z_t = U^\top Y_t, \quad t \in \mathbb{R}_+.$$

Then, (4.32) leads to

$$T^{-1} \int_0^T Z_t Z_t^\top\, ds \xrightarrow{a.s.} U^\top \Sigma U \text{ as } T \to \infty$$

under P^θ. Since $\int_0^T Z_t\, dW_t$ is a continuous P^θ-martingale, the martingale central limit theorem from Section 2.3.3 yields now

$$T^{-1/2}\sigma \int_0^T Z_t\, dW_t \xrightarrow{\mathcal{D}} N(0, \sigma^2 U^\top \Sigma U) \text{ under } P^\theta \text{ as } T \to \infty.$$

Since $Y_s = UZ_s$, the above convergence implies

$$T^{-1/2}\sigma \int_0^T Y_s\, dWs \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Sigma). \tag{4.33}$$

under P^θ as $T \to \infty$. Finally, we observe that

$$T^{1/2}\left(\hat{\theta}_T - \theta\right) = T^{-1/2} \tilde{V}_T T I_T^{-1}$$

and that by (4.32)

$$T^{-1} I_T \xrightarrow{a.s.} \Sigma \text{ as } T \to \infty$$

under P^θ such that the statement of the theorem follows from (4.33) and Slutsky's lemma. $\square$

5 Discrete observations: finite activity

In this chapter we consider the drift estimation problem for discrete observations. The maximum likelihood estimator for the drift given in (4.4) involves the continuous martingale part that is unknown when only discrete observations are given. Hence, we will approximate the continuous part of the process by removing observations that most likely contain jumps. We restrict our attention in this chapter to the case that the driving Lévy process has jumps of finite activity and use a threshold technique to distinguish increments of the process that contain jumps from increments that result from the continuous part only. This approach provides us in the high-frequency limit an asymptotically normal and efficient estimator. Based on these results we will treat the general case of an infinitely active jump component in Chapter 6.

In the literature on volatility estimation such jump filtering by thresholding was first employed by Mancini [2009] for estimating the integrated volatility of an Itô semimartingale. Aït-Sahalia and Jacod [2009] used similar methods to estimate the degree of activity of jumps in a general semimartingale setting.

5.1 Ornstein-Uhlenbeck type processes

5.1.1 Estimator and observation scheme

Let X be an Ornstein-Uhlenbeck process as in Section 3.2 and suppose we observe X at discrete time points $0 = t_1 < t_2 < \ldots < t_n = T_n$ such that $T_n \to \infty$ as well as $\Delta_n = \max_{1\leq i\leq n-1}\{|t_{i+1} - t_i|\} \downarrow 0$ and $n\Delta_n T_n^{-1} = O(1)$ as $n \to \infty$. The last condition assures that the number of observations n does not grow faster than $T_n\Delta_n^{-1}$ and it can always be fulfilled by neglecting observations, but it will simplify the formulation of the proof considerably. Denote by (b, σ^2, μ) the Lévy-Khintchine triplet of L. Assume throughout this section that $\lambda = \mu(\mathbb{R}) < \infty$ for the Lévy measure μ.

By deleting increments that are larger than some threshold $v_n > 0$ we filter increments that most likely contain jumps. This method applied to the estimator from Section 4.2 leads to the following estimator.

$$\bar{a}_n := -\frac{\sum_{i=0}^{n-1} X_{t_i}\Delta_i X \mathbf{1}_{\{|\Delta_i X|\leq v_n\}}}{\sum_{i=0}^{n-1} X_{t_i}^2 (t_{i+1} - t_i)}. \tag{5.1}$$

Here $v_n > 0, n \in \mathbb{N}$, is a cut-off sequence that will be chosen as a function of the maximal distance between observations and $\Delta_i X = X_{t_{i+1}} - X_{t_i}$.

In the finite activity case the jump part J of L can be written as a compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Z_i$$

where N is a Poisson process with intensity λ and the jump heights $Z_1, Z_2, \dots$ are iid with distribution F.

5.1.2 Asymptotic normality and efficiency

The indicator function that appears in $\bar{a}_n$ deletes increments that are larger than v_n. Lévy's modulus of continuity for the Wiener process implies that increments of the continuous part of X over an interval of length Δ_n are with high probability smaller than $\Delta_n^{1/2}$ (cf. Mancini [2009]). Hence, we set $v_n = \Delta_n^\beta$ for $\beta \in (0, 1/2)$ in order to keep the continuous part in the limit unaffected by the threshold. In order to be able to choose v_n such that $X_n^c = \sum_{i=0}^{n-1} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}}$ approximates the continuous martingale part in the limit, we make the following assumptions on the jumps of L and the observation scheme.

Assumption 5.1.1. (i) Suppose that there exists a $\beta \in (0, 1/2)$ such that the maximal distance between observation points satisfies $T_n \Delta_n^{(1-2\beta)\wedge\frac{1}{2}} = o(1)$,

(ii) the drift $b = 0$ and the distribution F of the jump heights is such that

$$F((-2\Delta_n^\beta, 2\Delta_n^\beta)) = o(T_n^{-1}).$$

Remark 5.1.2. Suppose that F has a bounded Lebesgue density f. Then $F((\Delta_n^\beta, \Delta_n^\beta)) = O(\Delta_n^\beta)$ and Assumption 5.1.1(ii) becomes $\Delta_n^\beta T_n = o(1)$. Together with 5.1.1(i) we obtain that $\beta = 1/3$ leads to an optimal compromise between 5.1.1(i) and (ii).

Remark 5.1.3. Assumption 5.1.1(i) means here that for given $T_n \to \infty$ we require $\Delta_n \downarrow 0$ fast enough such that there exists $\beta \in (0, 1/2)$:

$$T_n \Delta_n^{1-2\beta} = o(1) \text{ and } T_n \Delta_n^{1/2} = o(1).$$

Of course one of these to conditions will be dominating and determine the order of Δ_n.

Remark 5.1.4. Assumption 5.1.1(ii) gives a lower bound for the choice of the threshold β. At the same time 5.1.1(i) limits the range of possible β's from above, since the available frequency of observations, i.e. the order of Δ_n, may be limited in specific applications. Hence, the distribution F, the observation length T_n and frequency Δ_n fix a range for the choice of β. At this point the question of a data driven method to choose β arises, but this will not be considered in this work. The condition $b = 0$ is necessary in this context, since otherwise there is no hope to recover the continuous martingale part via jump filtering.

Remark 5.1.5. In contrast to the case of time-continuous observations in Chapter 4 the drift estimator (5.1) can be evaluated without any knowledge of the distribution of L provided that $b = 0$. More specifically the Lévy measure of L is not needed here anymore such that we consider in fact a semiparametric estimation problem.

The following theorem gives as the main result of this section a central limit theorem for the discretized MLE with jump filter.

Theorem 5.1.6. *Let X be stationary with finite second moments and assume that $\sigma^2 > 0$. Set $v_n = \Delta_n^\beta$ for $\beta \in (0, 1/2)$ and suppose that 5.1.1 holds, then*

$$T_n^{1/2}(\bar{a}_n - a) \xrightarrow{\mathcal{D}} N\left(0, \frac{\sigma^2}{E_a[X_0^2]}\right) \text{ as } n \to \infty.$$

The estimator $\bar{a}_n$ is asymptotically efficient.

Remark 5.1.7. We have proved in Chapter 4 that the MLE based on continuous observations attains the efficient asymptotic variance $\sigma^2/E_a[X_0^2]$. Since we cannot hope to obtain a lower asymptotic variance when only discrete observations are available, efficiency of $\bar{a}_n$ follows immediately from the first statement of Theorem 5.1.6.

5.1.3 Proof

We divide the proof of the theorem into several lemmas. First of all we need a probability bound for the event that the continuous component of X exceeds a certain threshold. Denote by W the Gaussian component of L and the drift part of X by

$$D_t = -a \int_0^t X_s \, ds.$$

Lemma 5.1.8. *Let $\sup_{s \in [0,T]}\{E[|X_s|^l]\} < \infty$ for some $l \geq 1$. For any $\delta \in (0, 1/2)$ and $i \in \{1, \ldots, n-1\}$ we have*

$$P(|\Delta_i W + \Delta_i D| > \Delta_n^{1/2-\delta}) = O\left(\Delta_n^{l(1/2+\delta)}\right) \text{ as } n \to \infty.$$

Proof: We may assume without loss of generality that $\sigma = 1$. In the first step we separate $\Delta_i W$ and $\Delta_i D$.

$$P\left(|\Delta_i W + \Delta_i D| > \Delta_n^{1/2-\delta}\right) \leq P\left(|\Delta_i W| > \frac{\Delta_n^{1/2-\delta}}{2}\right) + P\left(|\Delta_i D| > \frac{k\Delta_n^{1/2}}{2}\right).$$

By Lemma 22.2 in Klenke [2008]

$$P\left(|\Delta_i W| > \frac{\Delta_n^{1/2-\delta}}{2}\right) \leq 2\Delta_n^\delta \, e^{-\frac{1}{8\Delta_n^{2\delta}}}.$$

It follows from Jensen's inequality that

$$\left|\int_{t_i}^{t_{i+1}} X_s \, ds\right|^l \leq \Delta_n^{l-1} \int_{t_i}^{t_{i+1}} |X_s|^l \, ds.$$

This leads to

$$E[|\Delta_i D|^l] \leq \Delta_n^{l-1} a^l \int_{t_i}^{t_{i+1}} E[|X_s|^l]\, ds \leq \Delta_n^l a^l \sup_{s\in[t_i,t_{i+1}]} \{E[|X_s|^l]\}.$$

Finally, Markov's inequality yields

$$P\left(|\Delta_i D| > \frac{\Delta_n^{1/2-\delta}}{2}\right) \leq a^l \sup_{s\in[t_i,t_{i+1}]} \{E[|X_s|^l]\} \frac{2^l \Delta_n^l}{\Delta_n^{l(1/2-\delta)}} = O\left(\Delta_n^{l(1/2+\delta)}\right).$$

□

Jump filtering

First we will investigate how to choose the cut-off sequence v_n in order to filter the jumps. Define for $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$ the following sequence of events

$$A_n^i = \left\{\omega \in \Omega : \mathbf{1}_{\{|\Delta_i X| \leq v_n\}}(\omega) = \mathbf{1}_{\{\Delta_i N = 0\}}(\omega)\right\}.$$

Here N denotes the jump measure of L as defined in (2.4).

Lemma 5.1.9. *Suppose that Assumption 5.1.1 holds and set $v_n = \Delta_n^\beta, \beta \in (0, 1/2)$, then it follows that for $A_n = \bigcap_{i=1}^n A_n^i$ we have*

$$P(A_n) \to 1 \quad \textit{as } n \to \infty.$$

Proof: Observe that

$$P(A_n^c) = P\left(\bigcup_{i=1}^n (A_n^i)^c\right) \leq \sum_{i=1}^n P((A_n^i)^c).$$

By setting

$$K_n^i = \{|\Delta_i X| \leq v_n\},$$
$$M_n^i = \{\Delta_i N = 0\},$$

we can rewrite $(A_n^i)^c$ as

$$(A_n^i)^c = \left\{\mathbf{1}_{K_n^i} \neq \mathbf{1}_{M_n^i}\right\} = (K_n^i \backslash M_n^i) \cup (M_n^i \backslash K_n^i).$$

Here the events $K_n^i \backslash M_n^i$ and $M_n^i \backslash K_n^i$ correspond to the two types of errors that can occur when we search for jumps. In the first case we miss a jump and in the second case we neglect an increment although it does not contain any jumps. Next, we are going to bound the probability of both errors.

$$P((A_n^i)^c) = P(K_n^i \backslash M_n^i) + P(M_n^i \backslash K_n^i). \tag{5.2}$$

Set $\Delta_i = t_{i+1} - t_i$. For the first type of error we obtain

$$\begin{aligned} P(K_n^i \backslash M_n^i) =& P(|\Delta_i X| \le v_n, \Delta_i N > 0) \\ =& \sum_{j=1}^{\infty} e^{-\lambda \Delta_i} \frac{(\lambda \Delta_i)^j}{j!} P(|\Delta_i X| \le v_n | \Delta_i N = j) \\ \le& P(\Delta_i N = 1) P(|\Delta_i X| \le v_n | \Delta_i N = 1) + O(\Delta_n^2) \end{aligned} \tag{5.3}$$

and

$$\begin{aligned} P(|\Delta_i X| \le v_n | \Delta_i N = 1) \le& P(|\Delta_i X| \le v_n, |\Delta_i J| > 2v_n | \Delta_i N = 1) \\ &+ P(|\Delta_i X| \le v_n, |\Delta_i J| \le 2v_n | \Delta_i N = 1). \end{aligned} \tag{5.4}$$

The first term on the right side is bounded by

$$\begin{aligned} &P(|\Delta_i X| \le v_n, |\Delta_i J| > 2v_n | \Delta_i N = 1) \\ &= P(|\Delta_i W + \Delta_i J + \Delta_i D| \le v_n, |\Delta_i J| > 2v_n | \Delta_i N = 1) \\ &\le P(|\Delta_i W + \Delta_i D| > v_n, \Delta_i N = 1) P(\Delta_i N = 1)^{-1} \\ &\le P(|\Delta_i W + \Delta_i D| > v_n) P(\Delta_i N = 1)^{-1} = P(\Delta_i N = 1)^{-1} O(\Delta_n^{2-2\beta}), \end{aligned} \tag{5.5}$$

where we used Lemma 5.1.8. Denote by F the distribution of the jump heights of J. Then we obtain for the second term on the right-hand side of (5.4)

$$P(|\Delta_i X| \le v_n, |\Delta_i J| \le 2v_n | \Delta_i N = 1) \le P(|\Delta_i J| \le 2v_n | \Delta_i N = 1) = F((-2v_n, 2v_n)).$$

For the second addend in (5.2) it follows by independence of W and J that

$$\begin{aligned} P(M_n^i \backslash K_n^i) =& P(|\Delta_i X| > v_n, \Delta_i N = 0) \\ \le& P(|\Delta_i W + \Delta_i D| > v_n). \end{aligned}$$

Lemma 5.1.8 yields

$$P(|\Delta_i W + \Delta_i D| > v_n) = O(\Delta_n^{2-2\beta}). \tag{5.6}$$

Finally, (5.3), (5.5) and (5.6) lead to

$$P((A_n^i)^c) \le F((-2\Delta_n^{\beta}, 2\Delta_n^{\beta}))\Delta_n + O(\Delta_n^{2-2\beta})$$

such that the statement follows, since we have shown that

$$P(A_n^c) \le \sum_{i=1}^{n} P((A_n^i)^c) \le O(T_n) F((-2\Delta_n^{\beta}, 2\Delta_n^{\beta})) + O(T_n \Delta_n^{1-2\beta}).$$

$\square$

Approximation of the continuous martingale part

The crucial step in the proof is to show that the continuous martingale part can be approximated by thresholding. The following lemma does this step by proving that the numerator of the discretized MLE converges as its approximation with jump filter.

Lemma 5.1.10. *Under Assumption 5.1.1 we obtain*

$$\left|\sum_{i=0}^{n-1} X_{t_i}(\Delta_i X \mathbf{1}_{\{|\Delta_i X| \le v_n\}} - \Delta_i X^c)\right| = O_p(T_n \Delta_n^{1/2})$$

as $n \to \infty$.

Proof: On A_n from Lemma 5.1.9 we have

$$\sum_{i=0}^{n-1} X_{t_i} \left(\Delta_i X \mathbf{1}_{\{|\Delta_i X| \le v_n\}} - \Delta_i X^c\right) = \sum_{i=0}^{n-1} X_{t_i} \left(\Delta_i X \mathbf{1}_{\{\Delta_i N = 0\}} - \Delta_i X^c\right), \tag{5.7}$$

By Lemma 5.1.9 we have $P(A_n) \to 1$ as $n \to \infty$. Observe now that the difference of the increments on the right hand side of (5.7) is unequal to zero only if a jump occurred in that interval, i.e.

$$\Delta_i X \mathbf{1}_{\{\Delta_i N = 0\}} - \Delta_i X^c = \begin{cases} -\Delta_i X^c; & \Delta_i N > 0, \\ 0; & \Delta_i N = 0. \end{cases}$$

Define $C_i^n = \{\Delta_i N > 0\}$ and observe that

$$E_a \left| \mathbf{1}_{A_n} \sum_{i=0}^{n-1} X_{t_i}(\Delta_i X \mathbf{1}_{\{\Delta_i N = 0\}} - \Delta_i X^c)\right| = E_a \left|\sum_{i=0}^{n-1} X_{t_i} \Delta_i X^c \, \mathbf{1}_{A_n \cap C_i^n}\right|$$

The i-th increment of X^c can be written as $\Delta_i X = \Delta_i W + \Delta_i D$. Therefore,

$$\begin{aligned} E_a \left|\sum_{i=0}^{n-1} X_{t_i} \Delta_i X^c \, \mathbf{1}_{A_n \cap C_i^n}\right| &\le \sum_{i=0}^{n-1} E_a \left[|X_{t_i}(\Delta_i W + \Delta_i D)| \, \mathbf{1}_{A_n \cap C_i^n}\right] \\ &\le \sum_{i=0}^{n-1} E_a \left[(|X_{t_i} \Delta_i W| \, + |X_{t_i} \Delta_i D|) \, \mathbf{1}_{C_i^n}\right]. \end{aligned}$$

The number of jumps of J follows a Poisson process with intensity λ such that $P(C_i^n) \le \Delta_n \lambda$. The independence of $N \perp W$ and $\Delta_i N \perp X_{t_i}$ yields

$$\begin{aligned} \sum_{i=0}^{n-1} E_a \left[|X_{t_i} \Delta_i W| \ \ \mathbf{1}_{C_i^n}\right] = \sum_{i=0}^{n-1} E_a[(|X_{t_i}|] E[|\Delta_i W|] P(C_i^n) &\le E_a[|X_0|] T_n \Delta_n^{1/2} \lambda \\ &= O(T_n \Delta_n^{1/2}). \end{aligned}$$

Finally, by Hölder's inequality

$$\sum_{i=0}^{n-1} E_a \left[|X_{t_i} \Delta_i D| \ \mathbf{1}_{C_i^n} \right] \leq \sum_{i=0}^{n-1} E_a \left[X_{t_i}^2 (\Delta_i D)^2 \right]^{1/2} P(C_i^n)^{1/2} = O(T_n \Delta_n^{1/2}).$$

□

Central limit theorem for the discretized estimator

To prove Theorem 5.1.6, we need next that when we discretize the time-continuous estimator $\hat{a}_T$ from Section 4.2 as

$$\hat{a}_n = -\frac{\sum_{i=1}^n X_{t_i} \Delta_i X^c}{\sum_{i=1}^n X_{t_i}^2 (t_{i+1} - t_i)},$$

then $\hat{a}_n$ attains the same asymptotic distribution as $\hat{a}_T$ itself. This was already proved in Proposition 4.2.11. For completeness we give here the statement in the following lemma.

Lemma 5.1.11. *Let X be stationary with finite second moments and suppose that $\Delta_n T_n = o(1)$. Then*

$$T_n^{1/2} (\hat{a}_n - a) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2 E_a[X_0^2]^{-1}\right) \ \text{as } n \to \infty$$

under P^a.

In the last step we will then show that the discretized MLE and the estimator with jump filter show the same limiting behavior.

Proof of Theorem 5.1.6: By Lemma 5.1.11 $T_n^{1/2}(\hat{a}_n - a) \xrightarrow{\mathcal{D}} N(0, \frac{\sigma^2}{E_a[X_0^2]})$ as $n \to \infty$. By Slutsky's lemma it remains to show

$$T_n^{1/2} (\bar{a}_n - \hat{a}_n) \xrightarrow{p} 0 \text{ as } n \to \infty. \tag{5.8}$$

Observe that

$$T_n^{1/2} (\bar{a}_n - \hat{a}_n) = T_n^{1/2} \left(\frac{\sum_{i=1}^n X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}} - \sum_{i=1}^n X_{t_i} \Delta_i X^c}{\sum_{i=1}^n X_{t_i}^2 \Delta_i} \right).$$

By Lemma 5.1.10 we obtain under P_a that

$$T_n^{-1/2} \left(\sum_{i=1}^n X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}} - \sum_{i=1}^n X_{t_i} \Delta_i X^c \right) \xrightarrow{p} 0 \text{ as } n \to \infty,$$

and

$$T_n^{-1} \sum_{i=1}^n X_{t_i}^2 \Delta_i \xrightarrow{p} E_a[X_0^2],$$

such that (5.8) follows. □

5.2 Stochastic delay differential equations

In Section 4.5 we have derived a drift estimator for linear stochastic delay differential equations from time-continuous observations. Based on these results we will develop in this section an estimator for discrete observation and prove a central limit theorem for the estimation error. As for the Ornstein-Uhlenbeck process in the previous section we will approximate the continuous martingale part that appears in (4.29) by neglecting observations that are too large relative to the length of the time increment.

5.2.1 Model assumptions and drift estimator

Suppose that we observe the solution X of the following stochastic delay differential equation.

$$\begin{aligned} dX_t &= aX_t\, dt + cX_{t-1}\, dt + dL_t, && t > 0 \\ X_t &= X_t^0, && t \in [-1, 0], \end{aligned} \tag{5.9}$$

at discrete time points $0 = t_1 < \ldots < t_n = T_n$ where L is a Lévy process with Lévy-Khintchine triplet (b, σ^2, μ), drift parameter $\theta = (a, c) \in \mathbb{R}^2$ and $X^0 : [-1, 0] \times \Omega \to \mathbb{R}$ is the initial process with càdlàg trajectories that is assumed to be independent of L. Let the jump component J of L be of finite activity such that we can write J as a compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Z_i$$

where N is a Poisson process with intensity $\lambda = \mu(\mathbb{R})$ and the jump heights $Z_1, Z_2, \ldots$ are iid with distribution F.

To estimate the drift parameter θ we propose the following estimator based on (4.29)

$$\hat{\theta}_n = I_n^{-1} V_n, \tag{5.10}$$

where

$$V_n = \begin{pmatrix} \sum_{i=0}^{n-1} X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \le \Delta_n^\beta\}} \\ \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \le \Delta_n^\beta\}} \end{pmatrix}$$

and I_n is given by

$$I_n = \begin{pmatrix} \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i & \sum_{i=0}^{n-1} X_{t_i} X_{\lfloor t_i - 1 \rfloor_n} \Delta_i \\ \sum_{i=0}^{n-1} X_{t_i} X_{\lfloor t_i - 1 \rfloor_n} \Delta_i & \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n}^2 \Delta_i \end{pmatrix}.$$

Here, $\lfloor t_i - 1 \rfloor_n = \max_j \{t_j | t_j \le t_i - 1\}$ and $\beta \in (0, 1/2)$. The main result of this section is a central limit theorem for $\hat{\theta}_n$.

Theorem 5.2.1. *Suppose that Assumption 4.5.1 holds, $b = 0$ and that there exists $\beta \in (0, 1/2)$*

such that $T_n \Delta_n^{(1-2\beta)\wedge\frac{1}{2}} = o(1)$ *and* $F(-2\Delta_n^\beta, 2\Delta_n^\beta) = o(T_n^{-1})$*. Then*

$$T_n^{1/2}\left(\hat{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2\Sigma^{-1}\right)$$

where Σ *was defined in* (4.32).

5.2.2 Proof of Theorem 5.2.1

The proof of Theorem 5.2.1 is closely related to the asymptotic normality result for the Ornstein-Uhlenbeck drift estimator in Section 5.1. The main difference here is that due to the delay term we would need observations at $t_i - 1$, which are not available for irregularly spaced discrete observations. To overcome this problem we show that in the hight-frequency setting we are able to approximate X_{t_i-1} by the first observation $X_{\lfloor t_i-1\rfloor_n}$ before $t_i - 1$.

First of all we introduce some notation. Let W be the Gaussian component of L and denote the drift component of X by

$$D_t = a\int_0^t X_s\, ds + c\int_0^t X_{s-1}\, ds.$$

We define the following sequence of events to investigate the relation between the increments of X and the jumps of L. For $n \in \mathbb{N}$ and $i \in \{1, \ldots, n\}$ define

$$A_n^i = \left\{\omega \in \Omega : \mathbf{1}_{\{|\Delta_i X| \leq v_n\}}(\omega) = \mathbf{1}_{\{\Delta_i N = 0\}}(\omega)\right\}.$$

We divide now the proof of Theorem 5.2.1 into several lemmas. A close inspection of the proof of Lemma 5.1.9 shows that it has a straightforward extension to the delay setting such that for $A_n = \bigcap_{i=1}^n A_n^i$ we obtain

$$P(A_n) \to 1 \ \text{ as } n \to \infty. \tag{5.11}$$

Lemma 5.2.2. *If* $T_n\Delta_n = o(1)$ *we obtain*

$$\begin{aligned} T_n^{-1/2}\left|\sum_{i=0}^{n-1} X_{t_i}(\Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}} - \Delta_i X^c)\right| &= o_p(1), \\ T_n^{-1/2}\left|\sum_{i=0}^{n-1} X_{\lfloor t_i-1\rfloor_n}(\Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}} - \Delta_i X^c)\right| &= o_p(1), \end{aligned} \tag{5.12}$$

as $n \to \infty$.

Proof: First, we prove (5.12). Define $C_i^n = \{\Delta_i N > 0\}$. On A_n holds

$$
\begin{aligned}
&E\left|\mathbf{1}_{A_n}\sum_{i=0}^{n-1} X_{\lfloor t_i-1\rfloor_n}(\Delta_i X\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \Delta_i X^c)\right| \\
&= E\left|\mathbf{1}_{A_n}\sum_{i=0}^{n-1} X_{\lfloor t_i-1\rfloor_n}(\Delta_i X\mathbf{1}_{\{\Delta_i N=0\}} - \Delta_i X^c)\right| \\
&= E\left|\sum_{i=0}^{n-1} X_{\lfloor t_i-1\rfloor_n}\Delta_i X^c\mathbf{1}_{A_n\cap C_i^n}\right|
\end{aligned}
$$

The decomposition $X^c = W + D$ yields

$$
\begin{aligned}
E\left|\sum_{i=0}^{n-1} X_{\lfloor t_i-1\rfloor_n}\Delta_i X^c\mathbf{1}_{A_n\cap C_i^n}\right| &\le \sum_{i=0}^{n-1} E\left[|X_{\lfloor t_i-1\rfloor_n}\Delta_i W|\mathbf{1}_{C_i^n}\right] \\
&+ \sum_{i=0}^{n-1} E\left[|X_{\lfloor t_i-1\rfloor_n}\Delta_i D|\mathbf{1}_{C_i^n}\right]
\end{aligned}
$$

Now N is a Poisson process with intensity λ such that $P(C_i^n) \le \Delta_n\lambda$. The variables $X_{\lfloor t_i-1\rfloor_n}$, $\Delta_i W$ and $\Delta_i N$ are mutually independent such that

$$
\begin{aligned}
T_n^{-1/2}\sum_{i=0}^{n-1} E\left[|X_{\lfloor t_i-1\rfloor_n}\Delta_i W|\mathbf{1}_{C_i^n}\right] &\le T_n^{-1/2}\sum_{i=0}^{n-1} E\left[|X_{\lfloor t_i-1\rfloor_n}|\right] E\left[|\Delta_i W|\right] P(C_i^n) \\
&= O(T_n^{1/2}\Delta_n^{1/2}).
\end{aligned}
$$

Finally, Hölder's inequality and $X_{\lfloor t_i-1\rfloor_n} \perp C_i^n$ yields

$$
\begin{aligned}
&T_n^{-1/2}\sum_{i=0}^{n-1} E\left[|X_{\lfloor t_i-1\rfloor_n}\Delta_i D|\mathbf{1}_{C_i^n}\right] \\
&\le T_n^{-1/2}\sum_{i=0}^{n-1} E\left[\left(X_{\lfloor t_i-1\rfloor_n}\right)^2\right]^{1/2} E\left[(\Delta_i D)^2\right]^{1/2} P(C_i^n)^{1/2} = O(T_n^{1/2}\Delta_n^{1/2}).
\end{aligned}
$$

This completes the prove of (5.12). The second convergence follows by an analogous argument and we will skip the proof. □

In the next step we show that the pseudo estimator $\bar{\theta}_n$ defined by

$$\bar{\theta}_n = I_n^{-1}\bar{V}_n,$$

where

$$\bar{V}_n = \begin{pmatrix} \sum_{i=0}^{n-1} X_{t_i}\Delta_i X^c \\ \sum_{i=0}^{n-1} X_{\lfloor t_i-1\rfloor_n}\Delta_i X^c \end{pmatrix}$$

converges to the same asymptotic distribution as the estimator (4.29) based on time-continuous observations.

Lemma 5.2.3. *If $\Delta_n T_n = o(1)$, then*

$$T_n^{1/2}\left(\bar{\theta}_n - \theta\right) \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Sigma^{-1})$$

as $n \to \infty$.

The proof is similar to the Ornstein-Uhlenbeck case in Lemma 4.2.11 such that we will concentrate on the differences due to the bivariate parameter here.

Proof: Due to Girsanov's theorem the continuous P^0-martingale part can be written as

$$X_t^c = a \int_0^t X_s \, ds + b \int_0^t X_{s-1} \, ds + \sigma W_t = D_t + \sigma W_t,$$

where W is a P^θ-Wiener process. This leads to the decomposition

$$\bar{V}_n = \begin{pmatrix} \sum_{i=0}^{n-1} X_{t_i} \Delta_i D \\ \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n} \Delta_i D \end{pmatrix} + \sigma \begin{pmatrix} \sum_{i=0}^{n-1} X_{t_i} \Delta_i W \\ \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n} \Delta_i W \end{pmatrix} = \bar{V}_n^1 + \bar{V}_n^2 \tag{5.13}$$

and

$$T_n^{1/2}(\bar{\theta} - \theta) = T_n^{1/2}(I_n^{-1} \bar{V}_n^1 - \theta) + T_n I_n^{-1} T_n^{-1/2} \bar{V}_n^2.$$

Let us prove convergence of the second addend on the right hand side. The first step is to show that $T_n^{-1/2} \bar{V}_n^2 \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Sigma)$. This follows from (4.33), since Itô's isometry and Fubini's theorem yield for first entry of $(V_n^2 - \tilde{V}_{T_n})$ (recall the definition of $\tilde{V}_{T_n}$ in (4.31)) that

$$\begin{aligned} E\left[\left(\int_0^{T_n} X_{t-1} \, dW_t - \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n} \Delta_i W\right)^2\right] &= E\left[\left(\int_0^{T_n} \left(X_{t-1} - X_{\lfloor t-1 \rfloor_n}\right) \, dW_t\right)^2\right] \\ = \int_0^{T_n} E\left[\left(X_{t-1} - X_{\lfloor t-1 \rfloor_n}\right)^2\right] \, dt &= O(T_n \Delta_n). \end{aligned}$$

A similar estimate for the second entry yields $T_n^{-1/2}(V_n^2 - \tilde{V}_{T_n}) \xrightarrow{p} 0$ as $n \to \infty$. This together with (4.33) and Slutsky's lemma lead to

$$T_n I_n^{-1} T_n^{-1/2} \bar{V}_n^2 \xrightarrow{\mathcal{D}} N(0, \sigma^2 \Sigma^{-1}),$$

since (4.32) and a straightforward L^1-estimate for $T_n^{-1}(I_n - I_{T_n})$ shows that

$$T_n^{-1} I_n \xrightarrow{p} \Sigma \tag{5.14}$$

as $n \to \infty$. It remains to prove $T_n^{1/2}(I_n^{-1} \bar{V}_n^1 - \theta) \xrightarrow{p} 0$. Observe that

$$T_n^{1/2}(I_n^{-1} \bar{V}_n^1 - \theta) = T_n I_n^{-1} T_n^{1/2}(\bar{I}_n - I_n)\theta \tag{5.15}$$

such that the statement follows if $T_n^{-1/2}(\bar{I}_n - I_n) \xrightarrow{p} 0$, where we have rewritten $\bar{V}_n^1 = \bar{I}_n \theta$ using the matrix

$$\bar{I}_n = \begin{pmatrix} \sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} X_s \, ds & \sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} X_{s-1} \, ds \\ \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n} \int_{t_i}^{t_{i+1}} X_s \, ds & \sum_{i=0}^{n-1} X_{\lfloor t_i - 1 \rfloor_n} \int_{t_i}^{t_{i+1}} X_{s-1} \, ds \end{pmatrix}$$

An L^1-estimate for the first entry of $T_n^{-1/2}(\bar{I}_n - I_n)$ follows from Jensen's inequality:

$$\begin{aligned} &T_n^{-1/2} E\left[\left|\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i - \sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} X_s \, ds\right|\right] \\ &\leq \int_0^{T_n} E\left[\left|X_{\lfloor s \rfloor_n}^2 - X_s X_{\lfloor s \rfloor_n}\right|\right] \, ds = O(T_n^{1/2} \Delta_n^{1/2}). \end{aligned}$$

Three more similar estimates yield

$$T_n^{1/2}(\bar{I}_n - I_n) \xrightarrow{p} 0 \tag{5.16}$$

as $n \to \infty$ such that the statement follows from (5.14), (5.15) and (5.16). □

In the last step we show that $\hat{\theta}_n$ and $\bar{\theta}_n$ converge in distribution to the same limit.

Proof of Theorem 5.2.1: By Lemma 5.2.3

$$T_n^{1/2}(\bar{\theta}_n - \theta) \xrightarrow{\mathcal{D}} N(0, \Sigma^{-1}).$$

as $n \to \infty$. It remains to prove that $T_n^{1/2}(\hat{\theta}_n - \bar{\theta}_n) \xrightarrow{p} 0$ as $n \to \infty$. Observe that

$$T_n^{1/2}(\hat{\theta}_n - \bar{\theta}_n) = T_n I_n^{-1}(T_n^{-1/2}(V_n - \bar{V}_n)).$$

From Lemma 5.2.2 it follows that

$$T_n^{-1/2}(V_n - \bar{V}_n) \xrightarrow{p} 0$$

as $n \to \infty$ and $T_n^{-1} I_n \xrightarrow{p} \sigma^{-2}\Sigma$ such that finally by Slutsky's lemma

$$T_n^{1/2}(\hat{\theta}_n - \bar{\theta}_n) \xrightarrow{p} 0.$$

□

6 Discrete observations: infinite activity

In this chapter we discuss the estimation of the drift of an Ornstein-Uhlenbeck process from discrete observations when the jump part of the driving Lévy process can be of infinite activity. As in Chapter 5 our estimator will be based on deleting large increments in order to approximate the continuous martingale part of the process. We give conditions on the Lévy measure and suitable rates for the cut-off sequence that ensure separation in the high-frequency limit between jump part and continuous part. Under these conditions we will then prove asymptotic normality and efficiency of our method.

6.1 Estimator and observation scheme

Suppose an Ornstein-Uhlenbeck process as defined in (2.13) is observed at discrete, arbitrarily spaced time points $X_{t_0}, \dots, X_{t_n}$ for $0 = t_0 < \dots < t_n = T_n$. We denote the driving Lévy process by L and assume throughout this section that its drift vanishes such that the Lévy-Khintchine triplet is $(0, \sigma^2, \mu)$. As in chapter 5 we combine high-frequency and long-time asymptotics in order to approximate in the limit a continuous-time observation scheme, i.e.

$$\Delta_n = \max_{0 \le i \le n-1} \{|t_{i+1} - t_i|\} \downarrow 0 \text{ and } T_n \to \infty$$

as $n \to \infty$. Let us denote by $\Delta_i X = X_{t_{i+1}} - X_{t_i}$ the i-th increment of X. We consider the estimator from chapter 5 for the drift parameter a in the Ornstein-Uhlenbeck equation (2.13) given by

$$\bar{a}_n = -\frac{\sum_{i=0}^{n-1} X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \le v_n\}}}{\sum_{i=0}^{n-1} X_{t_i}(t_{i+1} - t_i)}.$$

In contrast to Chapter 5 we allow here also for Lévy processes with jumps of infinite activity. The sequence $v_n \in \mathbb{R}_+, n \in \mathbb{N}$ will be chosen such that increments larger than v_n most likely contain jumps. Results in Mancini [2009] imply that the size of the increments of X^c are of the order $\sqrt{\Delta_n \log 1/\Delta_n}$. We will choose $v_n = \Delta_n^\beta$ for $\beta \in (0, 1/2)$ such that in the limit $n \to \infty$ the continuous part of X is not affected by the threshold.

6.2 Asymptotic normality and efficiency

In this section we state as the main result of this chapter a CLT for the estimation error of $\bar{a}_n$. The limiting distribution will imply asymptotic efficiency of our method. But before we can formulate the theorem, we introduce some notation and assumptions on the jump part of L that enable us to separate the jump part and continuous part via jump filtering.

Let N denote the Poisson random measure associated to the jump part of L. The jump component J of X, the components M of jumps smaller than one and U of jumps larger than one and the drift D are given by

$$\begin{aligned} J_t &= \int_0^t \int_{-\infty}^{\infty} x(N(dx,ds) - \mu(dx)\lambda(ds)), \\ M_t &= \int_0^t \int_{-1}^{1} x(N(dx,ds) - \mu(dx)\lambda(ds)), \\ U_t &= J_t - M_t, \\ D_t &= -a \int_0^t X_s \, ds, \end{aligned} \tag{6.1}$$

respectively. Owing to this decomposition of X we can apply the results from Chapter 5 to D, W and U and thus can focus on M_t. To control the small jumps of M_t we impose the following assumption on the Lévy measure μ.

Assumption 6.2.1. (i) There exists an $\alpha \in (0,2)$ such that as $v \downarrow 0$

$$\int_{-v}^{v} x^2 \, \mu(dx) = O(v^{2-\alpha}). \tag{6.2}$$

(ii) There exists $\eta > 0$ such that for all $\epsilon \leq \eta$

$$E[\Delta_i M \mathbf{1}_{\{|\Delta_i M| \leq \epsilon\}}] = 0.$$

Remark 6.2.2. Assumption 6.2.1(i) controls the intensity of small jumps, which is determined by the mass of μ around the origin. When γ denotes the Blumenthal-Getoor index of L defined by

$$\gamma = \inf_{c \geq 0} \left\{ \int_{|x| \leq 1} |x|^c \, \mu(dx) < \infty \right\} \leq 2$$

then $\alpha = \gamma$ satisfies (6.2), i.e. Assumption 6.2.1(i) states that the Blumenthal-Getoor index is less than two. A similar conditions was used by Mancini [2011] in the context of volatility estimation. The second assumption is more of a technical nature and is needed to derive a moment bound for the small jump component M.

The main result of this chapter is the following central limit theorem for the drift estimator with jump filter.

Theorem 6.2.3. *Suppose that Assumption 4.2.4 and 6.2.1 hold and that X exhibits bounded second moments. If there exists $\beta \in (0,1/2)$ such that $T_n \Delta_n^{1-2\beta \wedge \frac{1}{2}} = o(1)$ as $n \to \infty$ then $v_n = \Delta_n^{\beta}$ yields*

$$T_n^{1/2}(\bar{a}_n - a) \xrightarrow{\mathcal{D}} N(0, \sigma^2 E_a[X_0^2]^{-1}).$$

The estimator is asymptotically efficient.

Remark 6.2.4. We have proved in Chapter 4 that the MLE based on continuous observations attains the efficient asymptotic variance $\sigma^2 E_a[X_0^2]^{-1}$. Since we cannot hope to obtain a lower asymptotic variance when only discrete observations are available, efficiency of $\bar{a}_n$ follows immediately from the first statement of Theorem 6.2.3.

Example 6.2.5. Let $L = W + J$, where J is a compound Poisson process

$$J_t = \sum_{i=1}^{N_t} Y_i$$

such that $Y_i \sim F$ are iid and N_t is a Poisson process with intensity λ. Suppose that F has a bounded Lebesgue density f. Then

$$\int_{-v}^{v} x^2 \ \mu(dx) = \lambda \int_{-v}^{v} x^2 f(x) \ dx \leq Cv^3$$

for $C > 0$ such that for L Assumption 6.2.1(i) holds for every $\alpha \in [0, 2)$.

More generally every Lévy process with Blumenthal-Getoor index less than two fulfills Assumption 6.2.1(i). This includes all Lévy processes commonly used in applications like (tempered) stable , normal inverse Gaussian, variance gamma and also gamma processes.

6.3 Proof

We will divide the proof of Theorem 6.2.3 into several lemmas. In the first place we prove a moment bound for the component of small jumps of L under jump filtering. This bound will be important in Lemma 6.3.8 to show that also the small jumps can be filtered in the limit. The next section is devoted to the problem of approximation of the continuous P^0-martingale part of X via jump filtering. Lemma 6.3.3 shows indeed that the integral with respect to X^c can be approximated in the high frequency limit by a thresholded version of X. In the proof of this lemma we make the decomposition (6.9) of the problem into three terms that reveal the structure of of the following parts of the proof. The first term is the approximation error of X^c by X^c plus compound Poisson jumps. This term has already been treated in Chapter 5. The second term measures the difference between thresholding with respect to X and with respect to X^c plus compound Poisson jumps and the last term contains the remaining component of small jumps.

In the rest of the proof we demonstrate that these three terms vanish in probability with rate $T_n^{-1/2}$. Lemma 6.3.4 verifies that the drift component of X is asymptotically not affected by the jump filter. Whereas Lemma 6.3.5 states that thresholding with respect to X and to the small jumps of L is asymptotically the same and in Lemma 6.3.7 we show that the continuous martingale part is not affected by thresholding X. Finally, we conclude the proof of Theorem 6.2.3 by applying Slutsky's lemma and the result of Lemma 6.3.3.

In the proofs in this section constants may change from line to line or even within one line without further notice.

6.3.1 A moment bound

In this section we derive a moment bound for short time increments of pure jump Lévy processes. Set

$$f(x) = \begin{cases} x^2 & ,\text{if } |x| \leq 1, \\ 0 & ,\text{if } |x| > 2 \end{cases}$$

and $f(x) \in [0,2]$ for $|x| \in (1,2]$ such that $f \in C^\infty(\mathbb{R})$. We scale f to be supported on $[-v,v]$ by

$$f^v(x) = v^2 f(x/v). \tag{6.3}$$

Proposition 6.3.1. *Let $(M_t)_{t\geq 0}$ be a pure jump Lévy process with Lévy measure μ such that* $\operatorname{supp}(\mu) \subset [-1,1]$ *and Assumption 6.2.1(1) and (2) hold. Then for all $\beta \in (0, \frac{1}{2})$ we obtain*

$$E\left[f^{t^\beta}(M_t)\right] = O\left(t^{1+\beta(2-\alpha)}\right)$$

as $t \downarrow 0$.

Remark 6.3.2. The estimate in Proposition 6.3.1 gives actually a bound for the Markov generator of M on the smooth test function f^v.

Proof: Let P^{M_t} denote the distribution of M_t. We apply Plancherel's identity to obtain

$$E[f^{t^\beta}(M_t)] = \int_{\mathbb{R}} f^{t^\beta}(x)\, P^{M_t}(dx) = (2\pi)^{-1} \int_{\mathbb{R}} \mathfrak{F} f^{t^\beta}(u) \overline{\phi_t(u)} du,$$

where $\mathfrak{F} f = \int_{\mathbb{R}} e^{iux} f(x)\, dx$ denotes the Fourier transform of f and the characteristic function of M satisfies

$$\phi_t(u) = \exp\left(t \int_{-1}^{1} (e^{iux} - 1 - iux)\, \mu(dx)\right).$$

Let us rewrite ϕ_t as the linearization of the exponential at zero plus a remainder R.

$$\phi_t(u) = 1 + \psi_t(u) + R(t,u)$$

with

$$\psi_t(u) = t \int_{-1}^{1} (e^{iux} - 1 - iux)\, \mu(dx).$$

Then,

$$\begin{aligned} E[f^{t^\beta}(M_t)] &= (2\pi)^{-1} \int_{\mathbb{R}} \mathfrak{F} f^{t^\beta}(u)(1 + \overline{\psi_t(u)} + \overline{R(t,u)})\, du \\ &= (2\pi)^{-1} \int_{\mathbb{R}} \mathfrak{F} f^{t^\beta}(u) \overline{\psi_t(u)}\, du + (2\pi)^{-1} \int_{\mathbb{R}} \mathfrak{F} f^{t^\beta}(u) \overline{R(t,u)}\, du. \end{aligned} \tag{6.4}$$

For the first term on the right hand side we obtain

$$
\begin{aligned}
&(2\pi)^{-1} \int_{\mathbb{R}} \mathfrak{F} f^{t^\beta}(u) \overline{\psi_t(u)}\, du = (2\pi)^{-1} t \int_{-1}^{1} \int_{\mathbb{R}} \mathfrak{F} f^{t^\beta}(u) (e^{-iux} - 1 + iux)\, du\, \mu(dx) \\
&= t \int_{-1}^{1} \left(f^{t^\beta}(x) + (2\pi)^{-1} \int_{\mathbb{R}} \mathfrak{F}\left((f^{t^\beta})' \right)(u) x\, du \right) \mu(dx) \\
&= t \int_{-1}^{1} f^{t^\beta}(x)\, \mu(dx) = tO(t^{\beta(2-\alpha)}) \qquad (6.5)
\end{aligned}
$$

by Assumption 6.2.1(i) and since

$$
\int_{\mathbb{R}} \mathfrak{F}\left(\left(f^{t^\beta} \right)' \right)(u)\, du = \left(f^{t^\beta} \right)'(0) = 0.
$$

It remains to bound the second addend in (6.4). For $Re(z) \leq 0$ observe that

$$
\left| \frac{e^z - z - 1}{z^2} \right| \leq C \qquad (6.6)
$$

for constant $C > 0$, since for $|z| \geq 1$

$$
\frac{|e^z - z - 1|}{z^2} \leq 2 + \frac{|z|}{z^2} \leq 3.
$$

Whereas on the half disk $\{|z| < 1, Re(z) \leq 0\}$ the continuous function $|e^z - z - 1|$ is bounded and z^2 is bounded except for the singularity at the origin, but at zero we know that $|e^z - z - 1| = O(z^2)$, i.e.

$$
\frac{|e^z - z - 1|}{z^2} \leq C < \infty
$$

on $\{|z| < 1, Re(z) \leq 0\}$. Theorem 2.2.5 in Kappus [2012] implies that $|\psi_t(u)| \leq Ct|u|^\alpha$ such that

$$
|R(t,u)| = |e^{\psi_t(u)} - \psi_t(u) - 1| \leq |\psi_t(u)|^2 \leq Ct^2 |u|^{2\alpha},
$$

where we used (6.6) and that for every characteristic function $|\exp(\psi_t(u))| = \phi_t(u) \leq 1$ holds. Hence, we obtain

$$
\left| \int_{\mathbb{R}} \mathfrak{F}\left(f^{t^\beta} \right)(u) \overline{R(t,u)}\, du \right| \leq Ct^2 \int_{\mathbb{R}} \left| \mathfrak{F}\left(f^{t^\beta} \right)(u) \right| |u|^{2\alpha}\, du. \qquad (6.7)
$$

Therefore, it remains to bound $\int_{\mathbb{R}} \left| \mathfrak{F} f^{t^\beta}(u) \right| |u|^{2\alpha}\, du$ in t. From (6.3) and the scaling property of the Fourier transform it follows that

$$
\mathfrak{F}(f^v)(u) = v^3 \mathfrak{F}\left(v^{-1} f(x/v) \right)(u) = v^3 \mathfrak{F}(f)(vu).
$$

Since $f \in C^\infty(\mathbb{R})$, we obtain $|\mathfrak{F}(f)(u)| \leq C_m |u|^{-m}$ such that

$$
|\mathfrak{F}(f^v)(u)| \leq C_m v^{3-m} |u|^{-m}
$$

for all $u \in \mathbb{R}$ and $m, v > 0$. Then

$$h(v,u) = |\mathfrak{F}(f^v)(u)||u|^{2\alpha} \leq C_m v^{3-m}|u|^{2\alpha-m}.$$

If

$$2\alpha + 1 < m \tag{6.8}$$

holds then $h(v,\cdot) \in L^1(\mathbb{R})$ for all $v \in (0,1)$. Setting $v = t^\beta$ yields

$$t^2 \int_{\mathbb{R}} \left|\mathfrak{F}\left(f^{t^\beta}\right)(u)\right| |u|^{2\alpha}\, du \leq C_m t^{(3-m)\beta+2}$$

for all $m > 0$. Since the first term in (6.4) is of the order $O\left(t^{1+\beta(2-\alpha)}\right)$, we choose m such that

$$(3-m)\beta + 2 \geq 1 + \beta(2-\alpha) \Leftrightarrow m \leq 1 + \beta^{-1} + \alpha.$$

Together with (6.8) this leads to the condition

$$2\alpha + 1 < 1 + \beta^{-1} + \alpha \Leftrightarrow \alpha < \beta^{-1},$$

which due to $\alpha \in (0,2)$ always holds for $\beta \in (0,1/2)$. Hence, we obtain

$$\left|\int_{\mathbb{R}} \mathfrak{F}\left(f^{t^\beta}\right)(u)\overline{R(t,u)}\, du\right| = O\left(t^{1+\beta(2-\alpha)}\right).$$

Together with (6.4) and (6.5) this yields finally

$$E[f^{t^\beta}(M_t)] = tO(t^{\beta(2-\alpha)}).$$

□

6.3.2 Approximating the continuous martingale part

The main step is to show that the continuous martingale part can be approximated by summing only the increments that are below the threshold v_n. We will use throughout the notation from (6.1).

Lemma 6.3.3. *Suppose that the assumptions of Theorem 6.2.3 hold, then*

$$T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i}(\Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}} - \Delta_i X^c) \xrightarrow{p} 0 \quad \textit{as } n \to \infty.$$

Proof: Let us consider the following decomposition where $\tilde{X} = W + D + U$

$$
\begin{aligned}
&T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}(\Delta_i X \mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \Delta_i X^c) = T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}(\Delta_i \tilde{X} \mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \Delta_i X^c) \\
&+ T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M \mathbf{1}_{\{|\Delta_i X|\le v_n\}} = T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}(\Delta_i \tilde{X} \mathbf{1}_{\{|\Delta_i \tilde{X}|\le 2v_n\}} - \Delta_i X^c) \\
&+ T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i \tilde{X}(\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \mathbf{1}_{\{|\Delta_i \tilde{X}|\le 2v_n\}}) + T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M \mathbf{1}_{\{|\Delta_i X|\le v_n\}} \\
&= S_1^n + S_2^n + S_3^n.
\end{aligned} \tag{6.9}
$$

The only difference between the term considered in Lemma 5.1.10 and S_1^n is that the drift $-a\int_0^t X_s ds$ in S_1^n integrates an OU process with possibly infinite jump activity. A careful analysis of the proof of Lemma 5.1.10 reveals that the same argument applies to S_1^n such that we conclude that S_1^n converges to zero in probability when $n \to \infty$. Let us prove next convergence of

$$
S_2^n = T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i \tilde{X}\left(-\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i \tilde{X}|\le 2v_n\}} + \mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i \tilde{X}|>2v_n\}}\right).
$$

First of all the contribution of the second indicator function on the right-hand side tends to zero in probability.

$$
\begin{aligned}
&P\left(T_n^{-1/2}\sum_{i=0}^{n-1} |X_{t_i}\Delta_i \tilde{X}| \mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i \tilde{X}|>2v_n\}} > 0\right) \\
&= P\left(\bigcup_{i=0}^{n-1}\{|\Delta_i X|\le v_n, |\Delta_i \tilde{X}|>2v_n\}\right) \le \sum_{i=0}^{n-1} P(|\Delta_i X|\le v_n, |\Delta_i \tilde{X}|>2v_n)
\end{aligned} \tag{6.10}
$$

When $|\Delta_i \tilde{X}| > 2v_n$ then with high probability $|\Delta_i U| > 0$, since by Lemma 5.1.8 we obtain

$$
\sum_{i=0}^{n-1} P(|\Delta_i \tilde{X}| > 2v_n, |\Delta_i U| = 0) \le \sum_{i=0}^{n-1} P(|\Delta_i W + \Delta_i D| > 2v_n) = O(T_n \Delta_n^{1-2\beta}). \tag{6.11}
$$

This together with (6.10) and the fact that on $\{|\Delta_i X| \le v_n, |\Delta_i \tilde{X}| > 2v_n\}$ necessarily $|\Delta_i M| > v_n$ implies that

$$
\begin{aligned}
&P\left(T_n^{-1/2}\sum_{i=0}^{n-1} |X_{t_i}\Delta_i \tilde{X}| \mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i \tilde{X}|>2v_n\}} > 0\right) \\
&\le \sum_{i=0}^{n-1} P(|\Delta_i U| \ne 0) P(|\Delta_i M| > v_n) + O(T_n \Delta_n^{1-2\beta}) \\
&= O(T_n \Delta_n v_n^{-2}) + O(T_n \Delta_n^{1-2\beta}) = O(T_n \Delta_n^{1-2\beta}),
\end{aligned}
$$

where we used Markov's inequality and independence of U and M. The remaining term in S_2^n is

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i\tilde{X}\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i\tilde{X}|\le 2v_n\}}.$$

Let us prove that on $\{|\Delta_i\tilde{X}|\le 2v_n\}$ the contribution of U is negligible.

$$\begin{aligned} & T_n^{-1/2}\sum_{i=0}^{n-1} P(|\Delta_i\tilde{X}|\le 2v_n, |\Delta_i U|>0) \\ &= T_n^{-1/2}\sum_{i=0}^{n-1}\Big(P(|\Delta_i\tilde{X}|\le 2v_n, \Delta_i N=1)+O(\Delta_n^2)\Big) \qquad (6.12)\\ &\le T_n^{-1/2}\sum_{i=0}^{n-1}\Big(P(|\Delta_i W+\Delta_i D|>1-2v_n)+O(\Delta_n^2)\Big) = O(T_n^{1/2}\Delta_n), \end{aligned}$$

as $n\to\infty$, where N denotes the counting process that counts the jumps of U and the last step follows from Lemma 5.1.8. Hence, we can assume that $\Delta_i U=0$ on $\{|\Delta_i\tilde{X}|\le 2v_n\}$ and so $\Delta_i\tilde{X}=\Delta_i W+\Delta_i D$. For $T_n\Delta_n^{1/2-\beta/2}=o(1)$ it follows from Lemma 6.3.4 that as $n\to\infty$

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i D\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i\tilde{X}|\le 2v_n\}} \xrightarrow{p} 0.$$

We have decomposed S_2^n into a term that converges to 0 in probability and a remainder.

$$S_2^n = T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i\tilde{X}|\le 2v_n\}} + o_p(1).$$

For the remainder let us observe that by Lemma 6.3.5 we obtain

$$\begin{aligned} S_2^n &= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i\tilde{X}|\le 2v_n\}} \\ &= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i W+\Delta_i D+\Delta_i M|>v_n, |\Delta_i W+\Delta_i D|\le 2v_n\}} + o_p(1) \\ &= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i M|>v_n\}} + o_p(1) \end{aligned}$$

Markov's inequality yields $P(|\Delta_i M|>v_n)\le\Delta_n^{1/2-\beta}$. Independence of X_{t_i}, $\Delta_i W$ and $\Delta_i M$ leads to

$$E\left[T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i M|>v_n\}}\right]=0$$

and

$$E\left[\left(T_n^{-1/2}\sum_{i=0}^{n-1}X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i M|>v_n\}}\right)^2\right]\leq T_n^{-1}E\left[\sum_{i=0}^{n-1}X_{t_i}^2(\Delta_i W)^2\mathbf{1}_{\{|\Delta_i M|>v_n\}}\right]$$

$$+T_n^{-1}E\left[\sum_{\substack{i,j\\ i\neq j}}X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i M|>v_n\}}X_{t_j}\Delta_j W\mathbf{1}_{\{|\Delta_j M|>v_n\}}\right].$$

Since $X_{t_i}, X_{t_j}, \Delta_i W, \Delta_i M, \Delta_j M \perp \Delta_j W$ for $i<j$ the off-diagonal elements are centered,

$$E\left[\sum_{\substack{i,j\\ i\neq j}}X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i M|>v_n\}}X_{t_j}\Delta_j W\mathbf{1}_{\{|\Delta_j M|>v_n\}}\right]=0$$

and the diagonal elements can be estimated by

$$T_n^{-1}E\left[\sum_{i=0}^{n-1}X_{t_i}^2(\Delta_i W)^2\mathbf{1}_{\{|\Delta_i M|>v_n\}}\right]\leq T_n^{-1}\Delta_n\sum_{i=0}^{n-1}E[X_{t_i}^2]P(|\Delta_i M|>v_n)$$

$$\leq\sup_i\{E[X_{t_i}^2]\}\Delta_n^{1/2-\beta}\to 0$$

as $n\to\infty$. The last step is to show that S_3^n tends to zero in probability as $n\to\infty$. As in (6.12) it follows that on $|\Delta_i X|\leq v_n$ we can assume that $\Delta_i U=0$. Now

$$\sum_{i=0}^{n-1}P(|\Delta_i X|\leq v_n,\Delta_i U=0)\leq\sum_{i=0}^{n-1}P(|\Delta_i W+\Delta_i D+\Delta_i M|\leq v_n,|\Delta_i M|\leq 2v_n)$$

$$+\sum_{i=0}^{n-1}P(|\Delta_i W+\Delta_i D+\Delta_i M|\leq v_n,|\Delta_i M|>2v_n).$$

The second addend vanishes, since by Lemma 5.1.8 we obtain

$$\sum_{i=0}^{n-1}P(|\Delta_i W+\Delta_i D+\Delta_i M|\leq v_n,|\Delta_i M|>2v_n)$$

$$\leq\sum_{i=0}^{n-1}P(|\Delta_i W+\Delta_i D|>v_n)=O(T_n\Delta_n^{1-2\beta}).$$

Thus, S_3^n can be rewritten as

$$S_3^n=T_n^{-1/2}\sum_{i=0}^{n-1}X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i X|\leq v_n,|\Delta_i M|\leq 2v_n\}}+o_p(1). \tag{6.13}$$

The convergence of the remaining term in S_3^n is dominated by the behavior of $\Delta_i M$ around the threshold, i.e. we prove next that

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i M|\le 2v_n\}} = T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}} + o_p(1).$$

Indeed,

$$\begin{aligned} &T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M(\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}} - \mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i M|\le 2v_n\}}) \\ &= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i X|> v_n, |\Delta_i M|\le 2v_n\}}. \end{aligned}$$

That last term tends to zero in probability will be shown in the proof of Lemma 6.3.5 below following equation (6.19). To finish the proof we demonstrate that the first addend on the right hand side of (6.13) vanishes asymptotically. Since X_{t_i} and $\Delta_i M$ are independent and $\Delta_i M$ is centered

$$E\left[T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}\right] = 0$$

and

$$\begin{aligned} E\left[\left(T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}\right)^2\right] &\le T_n^{-1}E\left[\sum_{i=0}^{n-1} X_{t_i}^2(\Delta_i M)^2\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}\right] \\ &+T_n^{-1}E\left[\sum_{\substack{i,j\\ i\neq j}} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}} X_{t_j}\Delta_j M\mathbf{1}_{\{|\Delta_j M|\le 2v_n\}}\right]. \end{aligned}$$

Since $X_{t_i}, X_{t_j}, \Delta_i M \perp \Delta_j M$ for $i<j$ the off-diagonal elements vanish by Assumption 6.2.1.1 such that

$$E\left[\sum_{\substack{i,j\\ i\neq j}} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}} X_{t_j}\Delta_j M\mathbf{1}_{\{|\Delta_j M|\le 2v_n\}}\right] = 0$$

and the diagonal elements can by Lemma 6.3.1 be estimated by

$$\begin{aligned} T_n^{-1}E\left[\sum_{i=0}^{n-1} X_{t_i}^2(\Delta_i M)^2\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}\right] &\le T_n^{-1}\sum_{i=0}^{n-1} E[X_{t_i}^2]E[(\Delta_i M)^2\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}] \\ &= \sup_i\{E[X_{t_i}^2]\}O(\Delta_n^{\beta(2-\alpha)}) \to 0 \end{aligned}$$

as $n\to\infty$. □

Approximation of the drift

The next step is to show that the drift component of X is in the limit not affected by the cut-off.

Lemma 6.3.4. *If the assumptions of Theorem 6.2.3 are fulfilled then*

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}(\Delta_i D\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \Delta_i D) \xrightarrow{p} 0 \quad \text{as } n\to\infty.$$

Proof: We rewrite the sum as follows.

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}(\Delta_i D\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \Delta_i D) = T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i D\mathbf{1}_{\{|\Delta_i X|> v_n\}}.$$

Next, we decompose $\Delta_i D$ as follows

$$\Delta_i D = -a\left(\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds + \Delta_i X_{t_i}\right)$$

such that by Lemma 6.3.5 below

$$\begin{aligned} T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i D\mathbf{1}_{\{|\Delta_i X|> v_n\}} =& -aT_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds\mathbf{1}_{\{|\Delta_i J|> v_n\}} \\ & - aT_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}^2\Delta_i\mathbf{1}_{\{|\Delta_i J|> v_n\}} + o_p(1). \end{aligned} \tag{6.14}$$

For the second term we obtain by Markov's inequality and from $v_n = \Delta_n^{\beta}$ that

$$E\left[\left|\sum_{i=0}^{n-1} X_{t_i}^2\Delta_i\mathbf{1}_{\{|\Delta_i J|> v_n\}}\right|\right] \le \sum_{i=0}^{n-1}\Delta_i E[X_{t_i}^2]P(|\Delta_i J|>v_n) \le CT_n\Delta_n^{1-2\beta}$$

and so for $T_n^{1/2}\Delta_n^{1-\beta} = o(1)$ it follows that

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\boldsymbol{\Delta}_i\mathbf{1}_{\{|\Delta_i J|> v_n\}} = o_p(1).$$

For the first sum on the right side of (6.14) we obtain by Hölder's inequality and independence of X_{t_i} and $\Delta_i J$

$$\begin{aligned} &E\left[\left|X_{t_i}\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds\mathbf{1}_{\{|\Delta_i J|> v_n\}}\right|\right] \le \\ E\left[\left(\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds\right)^2\right]^{1/2} &P(|\Delta_i J|>v_n)^{1/2}E[X_{t_i}^2]^{1/2} = O(\Delta_n^{3/2}v_n^{-1/2}) \end{aligned}$$

such that for $T_n^{1/2}\Delta_n^{1/2-\beta/2} = o(1)$ we can conclude that

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds\mathbf{1}_{\{|\Delta_i J|>v_n\}} = o_p(1).$$

□

6.3.3 Identifying the jumps

In the following we will show that the increments of X that are larger than the threshold v_n are dominated by the jump component.

Lemma 6.3.5.

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X(\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}) \xrightarrow{p} 0 \quad \textit{as } n\to\infty.$$

Proof: Observe that

$$\begin{aligned} T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X(\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \mathbf{1}_{\{|\Delta_i J|\le 2v_n\}}) &= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i J|>2v_n\}} \\ &\quad - T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i J|\le 2v_n\}}. \end{aligned} \tag{6.15}$$

We shall prove in Lemma 6.3.6 below that

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i J|>2v_n\}} \xrightarrow{p} 0. \tag{6.16}$$

In the next step we show that the contribution of U is negligible, since by independence of $\Delta_i W$, $\Delta_i M$, $\Delta_i U$ and X_{t_i} it follows that

$$\begin{aligned} E\left[\left|\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i J|\le 2v_n, |\Delta_i U|\neq 0\}}\right|\right] &\le \sum_{i=0}^{n-1} E[|X_{t_i}|]E[|\Delta_i W|]P(|\Delta_i U|\neq 0) \\ &\quad + \sum_{i=0}^{n-1} E[|X_{t_i}|]E[|\Delta_i J\mathbf{1}_{\{|\Delta_i U|\neq 0)\}}] \\ &\quad + \sum_{i=0}^{n-1} E[|X_{t_i}\Delta_i D|\mathbf{1}_{\{|\Delta_i U|\neq 0)\}}] \end{aligned} \tag{6.17}$$

Now U is a compound Poisson process with intensity $\mu(\mathbb{R}\backslash[-1,1]) < \infty$ such that $P(\Delta_i U \neq 0) = O(\Delta_n)$. We obtain for first summand on the right hand side

$$T_n^{-1/2} \sum_{i=0}^{n-1} E[|\Delta_i W|] E[|X_{t_i}|] P(\Delta_i U \neq 0) = O(T_n^{1/2} \Delta_n^{1/2}).$$

We split the second term into the contribution by U and J such that

$$\begin{aligned} T_n^{-1/2} \sum_{i=0}^{n-1} E[|X_{t_i}|] E[|\Delta_i J| \mathbf{1}_{\{\Delta_i U \neq 0\}}] &= T_n^{-1/2} \sum_{i=0}^{n-1} E[|X_{t_i}|] E[|\Delta_i M|] E[\mathbf{1}_{\{\Delta_i U \neq 0\}}] \\ &\quad + T_n^{-1/2} \sum_{i=0}^{n-1} E[|X_{t_i}|] E[|\Delta_i U| \mathbf{1}_{\{\Delta_i U \neq 0\}}] \end{aligned}$$

The first sum is of order

$$T_n^{-1/2} \sum_{i=0}^{n-1} E[|X_{t_i}|] E[|\Delta_i M|] E[\mathbf{1}_{\{\Delta_i U \neq 0\}}] = O(T_n^{1/2} \Delta_n^{1/2}).$$

Since U is a compound Poisson process, the sum in the second term is of the order of

$$N((t_i, t_{i+1}), \mathbb{R}\backslash[-1,1]).$$

This leads to the following estimate when n is large enough.

$$\begin{aligned} T_n^{-1/2} E \sum_{i=0}^{n-1} |X_{t_i} \Delta_i U \mathbf{1}_{\{\Delta_i U \neq 0\}}| &\leq T_n^{-1/2} E \sum_{i=1}^{N((0,T_n),\mathbb{R}\backslash[-1,1])} |X_{t_i} \Delta_i U \mathbf{1}_{\{\Delta_i U \neq 0\}}| \\ &= T_n^{-1/2} E[N((0,T_n), \mathbb{R}\backslash[-1,1])] E[|X_{t_i}|] E[\Delta_i U|] \\ &= O(T_n^{1/2} \Delta_n^{1/2}) \end{aligned}$$

where the sum up to N is over all increments that contain at least one jump of J. To prove convergence of the last addend in (6.17) we rewrite $\Delta_i D$ as follows

$$\Delta_i D = -a \left(\int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \, ds + \Delta_i X_{t_i} \right) \tag{6.18}$$

and so

$$\begin{aligned} T_n^{-1/2} E[| \sum_{i=0}^{n-1} X_{t_i} \Delta_i D \mathbf{1}_{\{\Delta_i U \neq 0\}}|] &\leq T_n^{-1/2} a E[| \sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i}) \, ds \mathbf{1}_{\{\Delta_i U \neq 0\}}|] \\ &\quad + T_n^{-1/2} E[| \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_n \mathbf{1}_{\{\Delta_i U \neq 0\}}|]. \end{aligned}$$

The first term on the right hand side gives by using Hölder's inequality

$$\begin{aligned}&T_n^{-1/2} aE\left[\left|\sum_{i=0}^{n-1} X_{t_i}\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds\mathbf{1}_{\{\Delta_i U\neq 0\}}\right|\right]\\&\leq T_n^{-1/2} a\sum_{i=0}^{n-1} E[X_{t_i}^2]^{1/2} E\left[\left(\int_{t_i}^{t_{i+1}}(X_s - X_{t_i})\,ds\right)^2\right]^{1/2}\mathbf{1}_{\{\Delta_i U\neq 0\}}|]\\&= O(T_n^{1/2}\Delta_n).\end{aligned}$$

Hence, we obtain

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i D\mathbf{1}_{\{\Delta_i U\neq 0\}} = O_p(T_n^{1/2}\Delta_n^{1/2})$$

such that it follows that

$$\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i J|\leq 2v_n, |\Delta_i U|\neq 0\}} = o_p(1).$$

Since the contribution of U is negligible, we obtain from (6.15) and (6.16) that

$$\begin{aligned}&T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X(\mathbf{1}_{\{|\Delta_i X|\leq v_n\}} - \mathbf{1}_{\{|\Delta_i M|\leq v_n\}})\\&= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i X|>v_n, |\Delta_i J|\leq 2v_n, |\Delta_i U|=0\}} + o_p(1).\end{aligned}$$

Hence, it remains to prove

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{\Delta_i X|>v_n|\Delta_i M|\leq 2v_n, |\Delta_i U|=0\}} \xrightarrow{p} 0 \quad \text{as } n\to\infty. \tag{6.19}$$

Observe that

$$\begin{aligned}&\{|\Delta_i M|\leq 2v_n, \Delta_i U = 0, |\Delta_i X| > v_n\}\\&\subset \{|\Delta_i M|\leq 2v_n, \Delta_i U = 0, |\Delta_i W + \Delta_i D| + |\Delta_i M| > v_n\}\\&\subset \{|\Delta_i W + \Delta_i D| > v_n/2\}\cup\{|\Delta_i M|\leq 2v_n, |\Delta_i M| > v_n/2\}.\end{aligned}$$

Therefore, the last two steps will be to show that

(i) $T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}} \xrightarrow{p} 0,$

(ii) $T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i M|>v_n/2, |\Delta_i M|\leq 2v_n\}} \xrightarrow{p} 0.$

For the proof of these two convergences we refer to Lemma 6.3.8 and Lemma 6.3.7. □

Lemma 6.3.6.

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X \mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i J|>2v_n\}} \xrightarrow{p} 0 \ \textit{as}\ n\to\infty$$

Proof: On $\{|\Delta_i X| \le v_n, |\Delta_i J| \ge 2v_n\}$ we have

$$||\Delta_i W + \Delta_i D| - |\Delta_i J|| \le |\Delta_i X| \le v_n.$$

Hence, we necessarily have $|\Delta_i W + \Delta_i D| > v_n$, i.e.

$$\{|\Delta_i X| \le v_n, |\Delta_i J| > 2v_n\} \subset \{|\Delta_i W + \Delta_i D| > v_n\} \tag{6.20}$$

such that

$$P(|\Delta_i X| \le v_n, |\Delta_i J| > 2v_n) \le P(|\Delta_i W + \Delta_i D| > v_n) = O(\Delta_n^{2-\beta}). \tag{6.21}$$

It follows from (6.20) that

$$\begin{aligned}
&T_n^{-1/2}\left|\sum_{i=0}^{n-1} X_{t_i}\Delta_i X \mathbf{1}_{\{|\Delta_i X|\le v_n, |\Delta_i J|>2v_n\}}\right| \le T_n^{-1/2}\sum_{i=0}^{n-1}|X_{t_i}\Delta_i X|\mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}}\\
&\le T_n^{-1/2}\sum_{i=0}^{n-1}|X_{t_i}\Delta_i W|\mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}} + T_n^{-1/2}\sum_{i=0}^{n-1}|X_{t_i}\Delta_i D|\mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}}\\
&+ T_n^{-1/2}\sum_{i=0}^{n-1}|X_{t_i}\Delta_i M|\mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}} + T_n^{-1/2}\sum_{i=0}^{n-1}|X_{t_i}\Delta_i U|\mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}}\\
&= A_n^1 + \ldots + A_n^4.
\end{aligned}$$

For A_n^1 we find by (6.21), Hölder's inequality and independence of X_{t_i} and $\Delta_i W$ that

$$\begin{aligned}
E[|A_n^1|] &\le T_n^{-1/2}\Delta_n^{1/2}\sum_{i=0}^{n-1} E[X_{t_i}^2]^{1/2} P(|\Delta_i W + \Delta_i D| > v_n)^{1/2}\\
&= O(T_n^{1/2}\Delta_n^{1/2-\beta/2}).
\end{aligned}$$

Using (6.18) we obtain for A_n^2 that

$$\begin{aligned}
E[|A_n^2|] \le\ & T_n^{-1/2} a \sum_{i=0}^{n-1} E\left[\left|X_{t_i}\int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds\right| \mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}}\right]\\
&+ T_n^{-1/2}\Delta_n \sum_{i=0}^{n-1} E[X_{t_i}^2 \mathbf{1}_{\{|\Delta_i W+\Delta_i D|>v_n\}}]
\end{aligned}$$

Hölder's inequality yields for the first term on the right hand side

$$
\begin{aligned}
&T_n^{-1/2} a \sum_{i=0}^{n-1} E\left[\left| X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds \right| \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right] \\
&\le T_n^{-1/2} a \sum_{i=0}^{n-1} E\left[\left(X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds \right)^2\right]^{1/2} P(|\Delta_i W + \Delta_i D| > v_n)^{1/2} \\
&= O(T_n^{1/2} \Delta_n^{1-\beta/2})
\end{aligned}
$$

for the second summand we find that

$$
\begin{aligned}
T_n^{-1/2} \Delta_n \sum_{i=0}^{n-1} E[X_{t_i}^2 \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}] &\le T_n^{-1/2} \Delta_n \sum_{i=0}^{n-1} E[X_{t_i}^4]^{1/2} P(|\Delta_i W + \Delta_i D| > v_n)^{1/2} \\
&= O(T_n^{1/2} \Delta_n^{1/2-\beta/2}).
\end{aligned}
$$

For A_n^3 we get by a similar estimate as for A_n^1 that

$$
E[|A_n^3|] = O(T_n^{1/2} \Delta_n^{1/2-\beta/2}).
$$

The last summand A_n^4 converges to zero in probability, since by independence and Hölder's inequality

$$
E[|A_n^4|] \le T_n^{-1/2} \sum_{i=0}^{n-1} E[X_{t_i}^2]^{1/2} E[\Delta_i U^2]^{1/2} P(|\Delta_i W + \Delta_i D| > v_n)^{1/2} = O(T_n^{1/2} \Delta_n^{1/2-\beta/2}).
$$

□

Now we show that the increments of the continuous part of X are negligible in the limit. This convergence is mainly based on the moment bound that we have derived in Lemma 5.1.8.

Lemma 6.3.7.

$$
T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}} \xrightarrow{p} 0 \ \text{as } n \to \infty.
$$

Proof: We decompose $\Delta_i X = \Delta_i W + \Delta_i D + \Delta_i M + \Delta_i U$ to obtain

$$
\begin{aligned}
&T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}} = T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i} \Delta_i W \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}} \\
&+ T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i} \Delta_i D \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}} + T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i} \Delta_i M \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}} \\
&+ T_n^{-1/2} \sum_{i=0}^{n-1} X_{t_i} \Delta_i U \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}} = V_n^1 + V_n^2 + V_n^3 + V_n^4.
\end{aligned}
$$

Lemma 5.1.8 yields for $k = \Delta_n^{\beta-1/2}$ and $l = 2$ that

$$P(|\Delta_i W + \Delta_i D| > v_n) = O(\Delta_n^{2-\beta}).$$

For V_n^1 we obtain by Hölder's inequality and independence of X_{t_i} and $\Delta_i W$ that

$$\begin{aligned}
E[|V_n^1|] &= T_n^{-1/2} E\left[\left|\sum_{i=0}^{n-1} X_{t_i} \Delta_i W \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right|\right] \\
&\leq T_n^{-1/2} \Delta_n^{1/2} \sum_{i=0}^{n-1} E[X_{t_i}^2]^{1/2} P(|\Delta_i W + \Delta_i D| > v_n) \\
&= O(T_n^{1/2} \Delta_n^{3/4-\beta}).
\end{aligned}$$

To prove convergence of V_n^2 we decompose $\Delta_i D$ as in (6.18) to obtain

$$\begin{aligned}
E[|V_n^2|] &= T_n^{-1/2} E\left[\left|\sum_{i=0}^{n-1} X_{t_i} \Delta_i D \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right|\right] \\
&\leq T_n^{-1/2} a E\left[\left|\sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right|\right] \\
&+ T_n^{-1/2} E\left[\left|\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right|\right]
\end{aligned}$$

Applying Hölder's inequality to the first term on the right hand side results in

$$\begin{aligned}
&T_n^{-1/2} a E\left[\left|\sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right|\right] \\
&\leq T_n^{-1/2} a \sum_{i=0}^{n-1} E\left[\left(X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds\right)^2\right]^{1/2} P(\Delta_i W + \Delta_i D| > v_n)^{1/2} \\
&= O(T_n^{1/2} \Delta_n^{1-\beta}).
\end{aligned}$$

The remaining term is of the order

$$\begin{aligned}
&T_n^{-1/2} E\left[\left|\sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \mathbf{1}_{\{|\Delta_i W + \Delta_i D| > v_n\}}\right|\right] \\
&\leq T_n^{-1/2} \Delta_n \sum_{i=0}^{n-1} E\left[X_{t_i}^4\right]^{1/2} P(|\Delta_i W + \Delta_i D| > v_n)^{1/2} \\
&= O(T_n^{1/2} \Delta_n^{3/4-\beta}).
\end{aligned}$$

Therefore, we conclude that $V_n^2 \to 0$ as $n \to \infty$. Similar estimates as for V_n^1 can be used for V_n^3 and V_n^4 to show

$$E[|V_n^3|] = O(T_n^{1/2}\Delta_n^{3/4-\beta})$$

and

$$E[|V_n^4|] = O(T_n^{1/2}\Delta_n^{3/4-\beta}).$$

This concludes the proof. □

The next lemma states that the increments of the jump component that are close to the threshold are negligible in the limit. For the proof we use the small time moment bound for the jump component from Lemma 6.3.1. This is the step where Assumption 6.2.1 on the intensity of small jumps becomes crucial.

Lemma 6.3.8.

$$T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{\frac{v_n}{2}<|\Delta_i M|\le 2v_n\}} \xrightarrow{p} 0 \;\; as\; n\to\infty.$$

Proof: Let us consider the following decomposition

$$\begin{aligned}
T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i M|>v_n/2,|\Delta_i M|\le 2v_n\}} &= T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i W\mathbf{1}_{\{|\Delta_i M|>v_n/2,|\Delta_i M|\le 2v_n\}}\\
+ T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i D\mathbf{1}_{\{|\Delta_i M|>v_n/2,|\Delta_i M|\le 2v_n\}} &+ T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i U\mathbf{1}_{\{|\Delta_i M|>v_n/2,|\Delta_i M|\le 2v_n\}}\\
+ T_n^{-1/2}\sum_{i=0}^{n-1} X_{t_i}\Delta_i M\mathbf{1}_{\{|\Delta_i M|>v_n/2,|\Delta_i M|\le 2v_n\}} &= S_n^1+S_n^2+S_n^3+S_n^4
\end{aligned}$$

For the probability that $|\Delta_i M|$ lies in $(v_n/2, 2v_n)$ we derive from Lemma 6.3.1 and Markov's inequality that

$$\begin{aligned}
P(|\Delta_i M|\le 2v_n, |\Delta_i M|>v_n/2) &= P(|\Delta_i M|\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}} > v_n/2)\\
&\le 4v_n^{-2}E[(\Delta_i M)^2\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}] = O(\Delta_n^{1-\alpha\beta}) \qquad (6.22)
\end{aligned}$$

Hence, by independence of X_{t_i}, $\Delta_i W$, and $\Delta_i M$ we find that $E[S_n^1] = 0$ and the second moment can be estimated as follows.

$$E[(S_n^1)^2] \leq T_n^{-1} E\left[\sum_{i=0}^{n-1} X_{t_i}^2 (\Delta_i W)^2 \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}}\right]$$
$$+ T_n^{-1} E\left[\sum_{\substack{i,j \\ i \neq j}} X_{t_i} \Delta_i W \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}} X_{t_j} \Delta_j W \mathbf{1}_{\{|\Delta_j M| > v_n/2, |\Delta_j M| \leq 2v_n\}}\right].$$

Since $X_{t_i}, X_{t_j}, \Delta_i W, \Delta_j M, \Delta_i M \perp \Delta_j W$ for $i < j$, the off-diagonal elements have zero expectation such that the second addend vanishes. For the diagonal elements we obtain

$$\begin{aligned} T_n^{-1} E\left[\sum_{i=0}^{n-1} X_{t_i}^2 (\Delta_i W)^2 \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}}\right] &\leq T_n^{-1} \Delta_n \sum_{i=0}^{n-1} E[X_{t_i}^2] O(\Delta_n^{1-\alpha\beta}) \\ &= O(\Delta_n^{1-\alpha\beta}) \end{aligned}$$

This yields the convergence $S_n^1 \xrightarrow{p} 0$ as $n \to \infty$. To prove that $S_n^2 \xrightarrow{p} 0$ as $n \to \infty$ we plug in (6.18) and obtain

$$\begin{aligned} E[|S_n^2|] = E&\left[a T_n^{-1/2} \left| \sum_{i=0}^{n-1} X_{t_i} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}} \right|\right] \\ &+ E\left[a T_n^{-1/2} \left| \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}} \right|\right] \end{aligned}$$

and by independence

$$\begin{aligned} &E\left[a T_n^{-1/2} \left| \sum_{i=0}^{n-1} X_{t_i}^2 \Delta_i \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}} \right|\right] \\ &\leq a T_n^{-1/2} \sum_{i=0}^{n-1} E[X_{t_i}^2] \Delta_i P(v_n/2 < |\Delta_i M| \leq 2v_n) \\ &= O(T_n^{1/2} \Delta_n^{1-\alpha\beta}). \end{aligned}$$

For the second term Hölder's inequality yields

$$E\left[a T_n^{-1/2} \left| \sum_{i=0}^{n-1} \int_{t_i}^{t_{i+1}} (X_s - X_{t_i})\, ds \mathbf{1}_{\{|\Delta_i M| > v_n/2, |\Delta_i M| \leq 2v_n\}} \right|\right] = O(T_n^{1/2} \Delta_n^{\frac{1-\alpha\beta}{2}})$$

From Assumption 6.2.1 it follows that S_n^4 is centered for n large enough. Furthermore, from Lemma 6.3.5 we conclude

$$\begin{aligned}E[(S_n^4)^2] &= T_n^{-1}\sum_{i=0}^{n-1} E[X_{t_i}^2]E[(\Delta_i M)^2\mathbf{1}_{\{|\Delta_i M|>v_n/2, |\Delta_i M|\le 2v_n\}}] \\ &\le T_n^{-1}\sum_{i=0}^{n-1} E[X_{t_i}^2]E[(\Delta_i M)^2\mathbf{1}_{\{|\Delta_i M|\le 2v_n\}}] \le O(\Delta_n^{(2-\alpha)\beta}) \overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

Finally, we show that $S_n^3 \overset{n\to\infty}{\longrightarrow} 0$. Independence together with (6.22) leads to

$$\begin{aligned}E[|S_n^3|] &= T_n^{-1/2}\sum_{i=0}^{n-1} E[|X_{t_i}\Delta_i U\mathbf{1}_{\{|\Delta_i M|>v_n/2, |\Delta_i M|\le 2v_n\}}|] \\ &= T_n^{-1/2}\sum_{i=0}^{n-1} E[|X_{t_i}|]E[|\Delta_i U|]P(|\Delta_i M|>v_n/2, |\Delta_i M|\le 2v_n) \\ &= O(T_n^{1/2}\Delta_n^{1-\alpha\beta})\end{aligned}$$

Proof of Theorem 6.2.3: Recall that for

$$\hat{a}_n = -\frac{\sum_{i=1}^n X_{t_i}\Delta_i X^c}{\sum_{i=1}^n X_{t_i}^2\Delta_i^n}$$

we already know that $T_n^{1/2}(\hat{a}_n - a) \xrightarrow{\mathcal{D}} N(0, \frac{\sigma^2}{E_a[X_0^2]})$ as $n\to\infty$. Therefore, it remains to show

$$T_n^{1/2}(\bar{a}_n - \hat{a}_n) \xrightarrow{p} 0 \text{ as } n\to\infty. \tag{6.23}$$

Observe that

$$T_n^{1/2}(\bar{a}_n - \hat{a}_n) = \frac{T_n^{-1/2}(\sum_{i=1}^n X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \sum_{i=1}^n X_{t_i}\Delta_i X^c)}{T_n^{-1}\sum_{i=1}^n X_{t_i}^2\Delta_i}.$$

By Lemma 6.3.3

$$T_n^{-1/2}\left(\sum_{i=1}^n X_{t_i}\Delta_i X\mathbf{1}_{\{|\Delta_i X|\le v_n\}} - \sum_{i=1}^n X_{t_i}\Delta_i X^c\right) \xrightarrow{p} 0 \text{ as } n\to\infty,$$

and

$$T_n^{-1}\sum_{i=1}^n X_{t_i}^2\Delta_i^n \xrightarrow{p} E_a[X_0^2],$$

such that (6.23) follows. □

7 Simulation results

In this chapter we investigate the finite sample performance of the estimators developed in Chapter 5 and 6 by means of Monte Carlo simulations. In the first section we consider Ornstein-Uhlenbeck type processes with finite jump intensity and give mean and standard deviation as well as the number of jumps detected for different parameter values and varying jump intensity. We also compare the distribution of the estimation error for finite samples and the asymptotic distribution from Theorem 5.1.6. The second part of this chapter is devoted to models with infinite jump activity.

7.1 Finite intensity models

In this section we perform Monte Carlo simulations for the drift estimator (5.1) of an Ornstein-Uhlenbeck type process defined by

$$X_t = e^{-at}X_0 + \int_0^t e^{-a(t-s)} dL_s, \quad t \in \mathbb{R}_+. \tag{7.1}$$

We take a deterministic starting value $X_0 \in \mathbb{R}$ and $a > 0$. The driving Lévy process L is assumed to be of the form

$$L_t = W_t + \sum_{i=1}^{N_t} Y_i,$$

where W is a Wiener process with $E[W_t^2] = \sigma_W^2 t$ and N is a Poisson process with intensity λ and the jump heights Y_i are iid with $N(0,2)$-distribution. An advantage of this Ornstein-Uhlenbeck model is that exact simulation algorithms are available both for X and L. We use an exact discretization of the explicit solution (7.1) to the Langevin equation driven by L on a equidistant time grid $t_i = \Delta_n i$ for $i = 1, \dots n$. Algorithms for the exact simulation of L can be found in Cont and Tankov [2004a] for example.

Table 7.1 contains means and standard deviations of each 100 realizations of the drift estimator $\bar{a}_n$ from (5.1). Since the Monte Carlo error is of order $N^{-1/2}$, where N is the number of Monte Carlo iterations, we have chosen a reasonable compromise between precision of the Monte Carlo approximation and computation time. The parameter values are $a = 2$ and 5 and jump intensity λ, time horizon T and number of observations n vary as given in the table. We also present the number of increments that were above the threshold $\Delta_n^{0.3}$. This number corresponds to the number of jumps that were detected and observe that it is extremely stable when T and λ are kept fixed, which suggests that the jump filter works quite reliable for finite intensity models and the threshold exponent $\beta = 0.3$. For the compound Poisson process the average number of jumps in an interval of length T is $E[N_T] = T\lambda$ is proportional to the jump intensity. This rule

			a = 2			a = 5		
λ	T	n	mean	std dev	∅ jumps detect	mean	std dev	∅ jumps detect
1	10	1000	2.0	0.3	6.7	5.0	0.5	7.4
		2000	2.0	0.3	7.0	5.0	0.5	7.2
		4000	2.0	0.4	7.0	5.0	0.5	6.8
	20	1000	2.0	0.2	13.1	4.7	0.3	12.5
		2000	2.0	0.2	13.2	4.9	0.4	12.3
		4000	2.0	0.2	13.0	5.0	0.3	13.1
	50	4000	2.0	0.1	31.3	4.8	0.2	31.2
		6000	2.0	0.2	32.2	4.6	0.3	30.1
5	10	1000	1.9	0.2	31.3	4.6	0.3	30.0
		2000	2.0	0.2	31.2	4.8	0.3	30.9
		4000	2.0	0.2	31.6	4.9	0.3	30.9
	20	2000	1.9	0.1	61.4	4.6	0.2	60.2
		4000	2.0	0.1	62.2	4.8	0.2	61.4
	50	4000	1.9	0.1	149	4.6	0.1	145
		6000	1.9	0.1	149	4.7	0.1	148

Table 7.1: Mean and standard deviation of $\bar{a}_n$ with $\beta = 0.3$ for an OU process with Gaussian component and compound Poisson jumps with intensity λ and the average number of increments filtered.

carries over to the jump filter with a surprising precision. We also see that the average number of filtered jumps is not equal to the expect number of jumps, but lies between 60 and 70 % of the former. This is surprising, since we would expect the average number of detected jumps to approach the expected number when Δ_n tends to zero.

Another interesting finding is that as soon as the step size Δ_n is so small that the discretization error is negligible (cf. Section 4.2.4 for an analysis of the discretization error) a further increase in the number of observations does not improve mean or standard deviation of the estimator any further. This indicates that the assumption of high-frequency observations is already reasonable when the stochastic error dominates the discretization error at least for finite intensity models.

The distribution of $T^{1/2}(\bar{a}_n - a)$ is shown in Figure 7.1 for $T = 70$ and $\Delta_n = 0.001$. The histogram on the left corresponds to $a = 2$ whereas on the right we have $a = 5$. From Theorem 5.1.6 and Lemma 2.4.4 it follows that the asymptotic variance of $\bar{a}_n$ is given by

$$\mathrm{AVAR}(\bar{a}_n) = (2a\sigma_W^2)\left(\sigma_W^2 + \lambda\sigma_j^2\right)^{-1}, \tag{7.2}$$

where σ_j^2 denotes the variance of the jump heights. Hence, we find that the asymptotic variance is proportional to a, which can also be observed for finite samples in Figure 7.1.

Figure 7.2 shows the same setup but with higher jump intensity. By comparing the results of Figure 7.1 and 7.2 we find that the variance also scales with the jump intensity as indicated in (7.2).

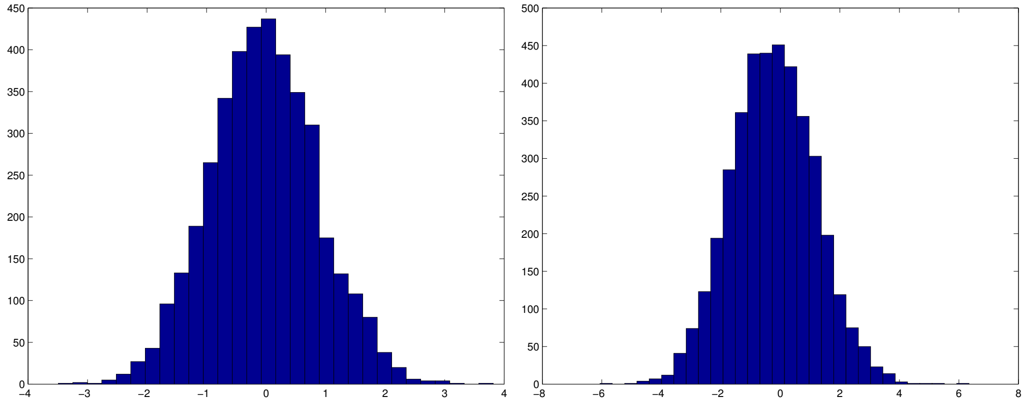

Figure 7.1: Error distribution of $\bar{a}_n$ for an Ornstein-Uhlenbeck process with $a = 2$ (left) and $a = 5$ (right), compound Poisson jumps with intensity $\lambda = 1$

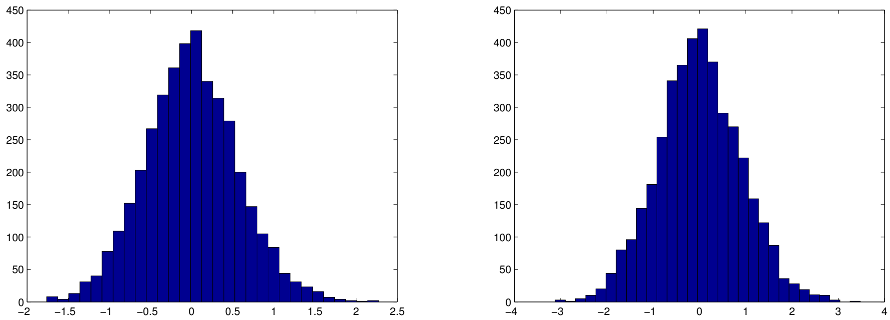

Figure 7.2: Error distribution of $\bar{a}_n$ for an Ornstein-Uhlenbeck process with $a = 2$ (left) and $a = 5$ (right), $\sigma_W = 1$, compound Poisson jumps with $\lambda = 3$

All in all we find that the estimator performs well even for very short time horizons if the discretization is fine enough. This observation corresponds to the results of Theorem 4.2.12 that states that the discretization bias is of the order $O(\Delta_n)$.

7.2 Infinite intensity models

In Chapter 6 we have proved an asymptotic normality result for the discretized maximum likelihood estimator with jump filter

$$\bar{a}_n = -\frac{\sum_{i=0}^{n-1} X_{t_i} \Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq v_n\}}}{\sum_{i=0}^{n-1} X_{t_i}(t_{i+1} - t_i)}.$$

also for models that involve a jump component of infinite activity. In this section we simulate data from an Ornstein-Uhlenbeck model of the form (7.1), where $L = W + G$ is a Wiener process W with $E[W_t^2] = \sigma_W^2 t$ and G is a gamma process. Again we consider an equidistant grid $t_i = i\Delta_n$ for $i = 0, \dots, n$. The gamma process has jumps of infinite activity, paths of finite variation and its Blumenthal-Getoor index is zero. The Lévy measure μ of G has an explicit Lebesgue density given by

$$g(x) = cx^{-1}e^{-\lambda x}\mathbf{1}_{\{x>0\}}$$

for $x \in \mathbb{R}$. The parameter $c > 0$ controls the jump intensity and $\lambda > 0$ the frequency of large jumps. It follows immediately from f that G is a subordinator. Exact simulation algorithms are known for increments of gamma processes and we use Johnk's algorithm (cf. Cont and Tankov [2004a]).

Table 7.2 gives mean and standard deviation for different observation length and parameter values. The 200 Monte Carlo iterations give a reasonable compromise between Monte Carlo error and computation time, which is of order 1/2. The standard deviation scales approximately with $T^{-1/2}$ as expected from Theorem 6.2.3. In contrast to Table 7.1 we kept here $\Delta_n = 0.0015$ fixed for all n. As in the finite intensity case we use the threshold exponent $\beta = 0.3$ for the jump filter. We find that the value of a has hardly any impact on the average number of increments that is filtered. When a increases the number of filtered increments also increases slightly, since a greater variability of the drift might push increments with a relatively small jump over the threshold.

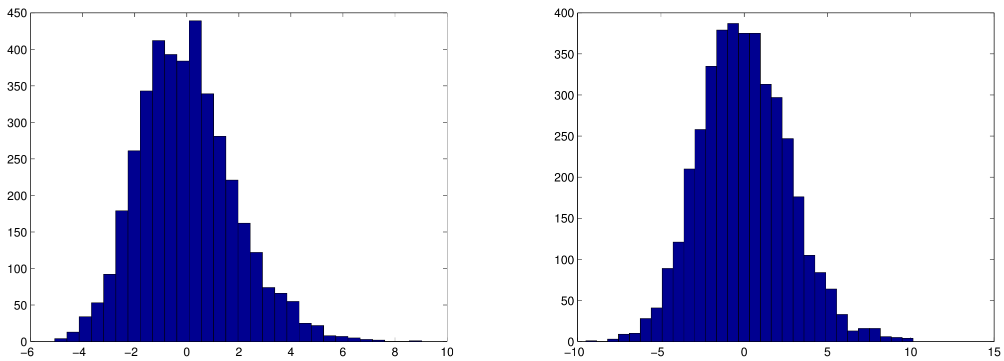

Figure 7.3: Error distribution of $\bar{a}_n$ for an Ornstein-Uhlenbeck process with $a = 2$ (left) and $a = 5$ (right), $\sigma_W = 1$ and Gamma process jumps

All in all we conclude that the jump filtering approach leads to good result also for models with infinite jump activity provided that the maximal observation distance is small.

7.3 Maximum likelihood vs. least squares estimation

In this section we compare maximum likelihood and least squares estimation for the Ornstein-Uhlenbeck type process X defined in (7.1). For continuously observed X the least squares

		a = 2			a = 5		
c	T	mean	std dev	∅ jumps detect	mean	std dev	∅ jumps detect
0.5	1	2.1	0.8	2.4	5.2	1.2	2.26
	5	2.0	0.4	11.7	5.0	0.6	12.1
	7.5	2.0	0.3	17.7	4.9	0.5	17.8
	10	2.0	0.3	23.73	5.0	0.4	23.9
	20	2.0	0.2	47.2	5.0	0.3	47.6
1	1	2.1	0.8	1.66	5.2	1.4	1.83
	2.5	2.1	0.6	5.2	5.1	1.1	5.90
	5	2.1	0.5	8.4	5.0	0.8	8.6
	7.5	2.0	0.5	12.7	5.0	0.7	13.1
	10	2.0	0.3	17.2	5.0	0.6	17.1

Table 7.2: Results of 200 Monte Carlo simulations of $\bar{a}_n$ with $\Delta_n = 0.0015$ and $\beta = 0.3$ for an Ornstein-Uhlenbeck process with $\sigma_W^2 = 1$ for the Gaussian component plus a Gamma process with $\lambda = 2$ and the average number of jumps filtered.

estimator for the drift parameter a is given by

$$\hat{a}_T^{LS} = -\frac{\int_0^T X_s dX_s}{\int_0^T X_s^2 ds}.$$

For Gaussian Ornstein-Uhlenbeck processes the least squares and the likelihood estimator $\hat{a}_T^{ML}$ (4.4) coincide, since the continuous martingale part under P^a equals the process itself. But when the driving process has jumps it follows from Theorem 4.2.10 that the asymptotic variances of both estimators are different by

$$\mathrm{AVAR}\left(\hat{a}_T^{LS}\right) - \mathrm{AVAR}\left(\hat{a}_T^{ML}\right) = E_a[X_0^2]^{-1} \int_{\mathbb{R}} x^2 \, \mu(dx) > 0.$$

Hence, the least squares estimator is inefficient when jumps are present.

Figure 7.4 compares the mean of the MLE and LSE for compound Poisson jumps with different jump intensities. For each intensity the mean of 500 Monte Carlo simulations is given for an Ornstein-Uhlenbeck process with $\sigma_W = 1$ and jumps with $N(0,2)$-distribution. The true parameter is $a = 2$ and we find that the MLE performs slightly better than the LSE.

In Figure 7.5 the standard deviation for both estimators is given. The jump intensity λ of the compound Poisson part of L varies between one and ten. In this model setup the difference in asymptotic variance between MLE and LSE is given by

$$\mathrm{AVAR}\left(\hat{a}_T^{LS}\right) - \mathrm{AVAR}\left(\hat{a}_T^{ML}\right) = \frac{2a\sigma_j^2\lambda}{\sigma_W^2 + \sigma_j^2\lambda}$$

We observe that already for small jump intensities the MLE clearly outperforms the LSE. With growing intensity this efficiency gain becomes even more severe. For $\lambda = 10$ the standard

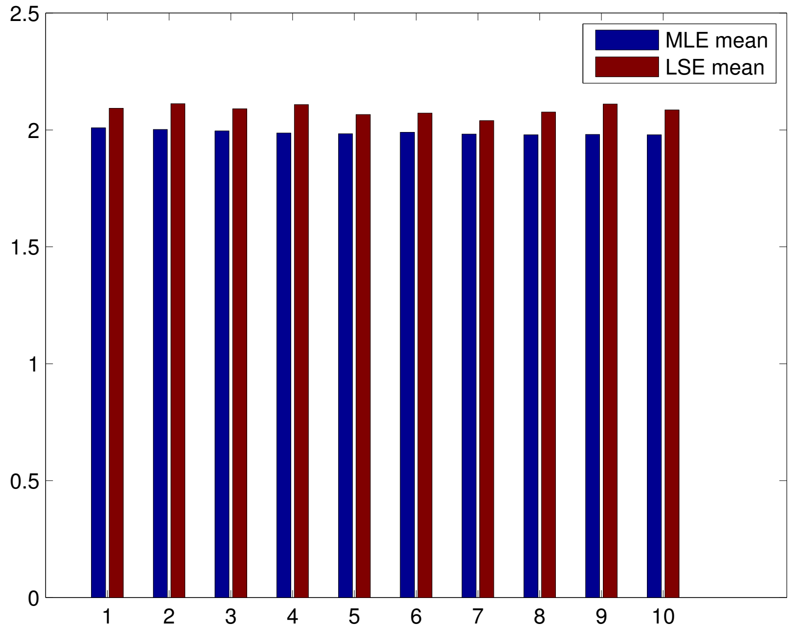

Figure 7.4: Mean of MLE (blue) and LSE (red) for varying jump intensity

deviation is about five times larger for the least squares estimator.

This short simulation example shows the significant gain in efficiency when we use a discretized likelihood estimator with approximation of the continuous part for drift estimation and underlines the importance of jump filtering for jump diffusion models.

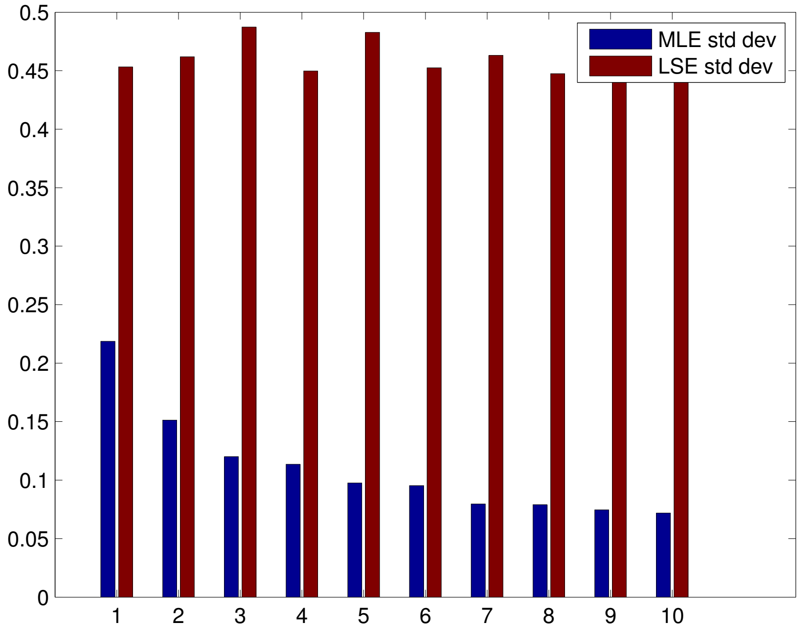

Figure 7.5: Standard deviation of MLE (blue) and LSE (red) for varying jump intensity

8 Conclusion

The goal of this thesis was to develop explicit maximum likelihood estimators for the drift of Lévy-driven jump diffusion processes that are efficient in the Hájek-Le Cam sense. This rather classical problem of parametric drift estimation has led to some new mathematical challenges. Separation of the continuous and the jump part of a process under high-frequency observations is currently a central topic for the statistical analysis of semimartingale models (cf. e.g. Jacod and Protter [2012] and Fan and Wang [2007]) and in the theoretical analysis of this field many questions are still open. To the best of our knowledge this is the first work that considers the problem of recovering the continuous martingale part under high-frequency asymptotics $\Delta_n \downarrow 0$ together with a growing observation horizon, i.e. $T_n \to \infty$, in the context of jump filtering.

In Chapter 5 and 6 we have demonstrated that the likelihood theory for continuous observations from Chapter 3 and 4 leads to efficient drift estimators when jump filtering is applied for discrete observations with growing time horizon. We would like to emphasize that this estimation approach applies also to a much wider setting. Let us sketch some possible extensions:

- **Jump diffusions with affine drift parameter**
  The jump diffusions X with affine drift parameter (cf. Section 4.4) was defined as a strong solution to

$$\begin{aligned} dX_t &= (g(t, X_t) + \theta f(t, X_t))\ dt + \gamma(t, X_t)\ dL_t, \quad t \geq 0 \\ X_0 &= x \in \mathbb{R} \end{aligned}$$

  A maximum likelihood estimator for θ based on time-continuous observations was derived in Section 4.4. Suppose arbitrarily spaced observations $X_{t_1}, \dots, X_{t_n}$ are given. Approximation of the continuous martingale part via jump filtering provides the following estimator for θ:

$$\hat{\theta}_n = \frac{\sum_{i=0}^{n-1} \gamma(t_i, X_{t_i})^2 f(t_i, X_{t_i}) \Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq \Delta_n^\beta\}}}{\sum_{i=0}^{n-1} \gamma(t_i, X_{t_i})^2 f(t_i, X_{t_i})(t_{i+1} - t_i)}$$

  for $\beta \in (0, 1/2)$. To derive asymptotic properties of $\hat{\theta}_n$ the methods from Chapter 5 and 6 can be generalized under similar conditions on the Lévy measure μ of L and the observation scheme.

- **Numerical evaluation of the MLE for general drift coefficients**
  We can even go one step further. Also for the general jump diffusion model (3.1) jump filtering plays an important role. In this setting the maximum likelihood estimator was defined in (4.2) and in this generality it can only be evaluated numerically. But before numerical procedures can be applied the continuous martingale part in the likelihood

function

$$\mathcal{L}(X,\theta)_T = \exp\left(\int_0^T c(s,X_s)^{-1}\delta(\theta,s,X_s)\, dX_s^c \right.$$
$$\left. - \frac{1}{2}\int_0^T \delta(\theta,s,X_s)^\top c(s,X_s)^{-1}\delta(\theta,s,X_s)\, ds\right),$$

has to be approximated. This leads to the estimator

$$\hat{\theta}_n = \arg\max_{\theta\in\Theta} \exp\left(\sum_{i=0}^{n-1}\gamma(t_i,X_{t_i})^{-2}\delta(\theta,t_i,X_{t_i})\Delta_i X \mathbf{1}_{\{|\Delta_i X|\le \Delta_n^\beta\}} \right.$$
$$\left. -\frac{1}{2}\sum_{i=0}^{n-1}\gamma(t_i,X_{t_i})^{-2}\delta\left(\theta,t_i,X_{t_i}^2\right)(t_{i+1}-t_i)\right)$$

for $\beta \in (0,1/2)$ that can now be computed numerically.

- **The pure jump case** $\sigma^2 = 0$
  In this work we have considered the case of a non-trivial Gaussian component of L (i.e. $\sigma^2 > 0$), since this is a necessary condition for local equivalence of the induced measures and thus for the existence of the likelihood function. But also for $\sigma^2 = 0$ jump filtering can be used to infer the drift parameter.

  When L is a pure jump Lévy process with Lévy-Khintchine triplet $(0,0,\mu)$ and X is the jump diffusion defined by

  $$\begin{aligned} dX_t &= (g(t,X_t) + \theta f(t,X_t))\ dt + \gamma(t,X_t)\ dL_t, \quad t\ge 0\\ X_0 &= x \in \mathbb{R}\end{aligned}$$

  we can define the pseudo estimator

  $$\hat{\theta}_n = \frac{\int_0^T \gamma(t,X_t)^{-2} f(t,X_t)\ dX_t^c}{\int_0^T \gamma(t,X_t)^{-2} f(t,X_t)\ dt}. \tag{8.1}$$

  Then at least formally the continuous P^θ-martingale part is given by

  $$X_t^c = \theta\int_0^t f(x,X_s)\ ds$$

  such that $\hat{\theta}_T \overset{a.s.}{=} \theta$ under P^θ. Hence, for full observations $(X_t)_{t\in[0,T]}$ the drift parameter is known even for finite time horizon $T<\infty$. This is in direct contrast to the case $\sigma^2 > 0$, where the drift can only be recovered when $T\to\infty$.

  For discrete observations $X_{t_1},\dots,X_{t_n}$ these considerations imply that a high frequency scheme $\Delta_n \downarrow 0$ for $0 = t_1 < \ldots < t_n = T$ and $T>0$ fixed is appropriate. By jump filtering we obtain the estimator

$$\hat{\theta}_n = \frac{\sum_{i=0}^{n-1} \gamma(t_i, X_{t_i})^2 f(t_i, X_{t_i}) \Delta_i X \mathbf{1}_{\{|\Delta_i X| \leq \Delta_n^\beta\}}}{\sum_{i=0}^{n-1} \gamma(t_i, X_{t_i})^2 f(t_i, X_{t_i})(t_{i+1} - t_i)}$$

for $\beta \in (0, 1/2)$. Laws of larger numbers and central limit theorems for functionals of the form

$$V^n(f, v_n, X)_t = \sum_{i=0}^{n-1} f(\Delta_i X) \mathbf{1}_{\{|\Delta_i X| \leq v_n\}}$$

for a threshold $v_n = C\Delta_n^\beta$, $\beta \in (0, 1/2)$, and polynomial growth conditions on the test function f were shown in Jacod and Protter [2012] for $\Delta_n \downarrow 0$ and fixed $T > 0$ under the assumption that $\sigma^2 > 0$. Convergence results in this setting for $\sigma^2 = 0$ are still an open problem.

All these extensions show that the scope of the methods in this thesis is much wider than the models that we have investigated in detail in Chapter 5 and 6. Our main goal here was to understand the limits of the jump filtering approach when different jump behaviors of the driving Lévy process are considered. In Chapter 6 we found that it applies in the infinite activity case under mild conditions on the small jumps. The main restriction (cf. Assumption 6.2.1(i)) is that the Blumenthal-Getoor index is strictly less than two is necessary for the approximation of the continuous martingale part. When the Blumenthal-Getoor index is approaching two the jump part of L behaves more and more like a Wiener process such that the separation of those two becomes more and more difficult. Hence, this is a natural condition in the context of jump filtering that is necessary for the identifiability of the continuous part (see also Cont and Mancini [2011]).

Computational efficiency is another advantage of the likelihood approach that leads in many popular models to explicit estimators that are in contrast to other techniques for jump diffusions extremely easy to implement. This is a result of starting from the time-continuous likelihood function, which is explicitly known such that no Monte Carlo approximations of unknown transition densities or moments are needed. The same holds true for the jump filtering technique.

To draw a conclusion maximum likelihood techniques for jump diffusion processes lead to efficient and easy to implement drift estimators that are also computationally efficient. The jump filtering that is needed here to extract the continuous component that contains information on the drift applies under very mild conditions on the jump behavior of the driving Lévy process and makes the likelihood approach feasible for discretely observed processes.

Bibliography

Yacine Aït-Sahalia and Jean Jacod. Volatility estimators for discretely sampled Lévy processes. *Ann. Stat.*, 35(1):355–392, 2007.

Yacine Aït-Sahalia and Jean Jacod. Estimating the degree of activity of jumps in high frequency data. *Ann. Stat.*, 2009.

Yacine Aït-Sahalia and Jean Jacod. Identifying the successive Blumenthal-Getoor indices of a discretely observed process. *Ann. Stat.*, 2012.

David Applebaum. *Lévy processes and stochastic calculus. 2nd ed.* Cambridge Studies in Advanced Mathematics 116. Cambridge University Press., 2009.

Federico M. Bandi and Thong H. Nguyen. On the functional estimation of jump-diffusion models. *J. Econom.*, 116(1-2):293–328, 2003.

Ole E. Barndorff-Nielsen and Neil Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial economics. *J. R. Stat. Soc., Ser. B, Stat. Methodol.*, 63(2): 167–241, 2001.

David S. Bates. Jumps and stochastic volatility: the exchange rate processes implicit in deutschemark options. *Rev. Fin. Studies*, 9:69–107, 1996.

Denis Belomestny and John Schoenmakers. A jump-diffusion Libor model and its robust calibration. *Quant. Finance*, 11(4):529–546, 2011.

Jean Bertoin. *Lévy processes.* Cambridge University Press, 1998.

Markus Bibinger. Efficient covariance estimation for asynchronous noisy high-frequency data. *Scandinavian Journal of Statistics*, 38:23–45, 2011.

David Blackwell. Comparison of experiments. Proc. Berkeley Sympos. math. Statist. Probability, California July 31 - August 12, 1950, 93-102., 1951.

Peter J. Brockwell, Richard A. Davis, and Yu Yang. Estimation for nonnegative Lévy-driven Ornstein-Uhlenbeck processes. *J. Appl. Probab.*, 44(4):977–989, 2007.

Emmanuelle Clement, Sylvain Delattre, and Arnaud Gloter. Optimality properties in estimating jumps. preprint. available at http://hal.archives-ouvertes.fr/hal-00609983. 2011.

Rama Cont and Cecilia Mancini. Nonparametric tests for pathwise properties of semimartingales. *Bernoulli*, 17(2):781–813, 2011.

Rama Cont and Peter Tankov. *Financial modelling with jump processes.* Chapman & Hall, 2004a.

Rama Cont and Peter Tankov. Non-parametric calibration of jump-diffusion option pricing models. *J. Comput. Finance*, 7, 2004b.

John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. A theory of the term structure of interest rates. *Econometrica*, (53):363–384, 1985.

Anirban DasGupta. *Asymptotic theory of statistics and probability.* Springer Texts in Statistics. New York: Springer , 2008.

Donald A. Dawson and Zenghu Li. Skew convolution semigroups and affine Markov processes. *Ann. Probab.*, 34(3):1103–1142, 2006.

Claude Dellacherie and Paul-Andre Meyer. *Probabilités et potentiel. Chapitres V à VIII: Théorie des martingales. Ed. ent. ref.* Actualites scientifiques et industrielles, 1385. Publications de l'Institut de Mathematique de l'Universite de Strasbourg, XVII. Paris: Hermann., 1980.

Paul Doukhan. *Mixing: Properties and examples.* Lecture Notes in Statistics 85. New York: Springer, 1994.

Richard M. Dudley. *Real analysis and probability. Repr.* Cambridge Studies in Advanced Mathematics. 74. Cambridge University Press., 2002.

Richard M. Dudley and R. Norvaisa. *Product integrals, Young integrals and p-variation..* Lecture Notes in Mathematics. Springer, New York, 1998.

Darrell Duffie, Jun Pan, and Kenneth Singleton. Transform analysis and asset pricing for affine jump-diffusions. *Econometrica*, 68(6):1343–1376, 2000.

Darrell Duffie, Jun Pan, and Kenneth Singleton. Affine processes and applications in finance. *Ann. Appl. Probab.*, 13(3):984–1053, 2003.

Jianqing Fan and Yazhen Wang. Multi-scale jump and volatility analysis for high-frequency financial data. *J. Am. Stat. Assoc.*, 102(480):1349–1362, 2007.

Paul David Feigin. Maximum likelihood estimation for continuous-time stochastic processes. *Adv. Appl. Probab.*, 8:712–736, 1976.

William Feller. Two singular diffusion problems. *Ann. Math. (2)*, 54:173–182, 1951.

Willy Feller. On the integro-differential equations of purely discontinuous Markoff processes. *Trans. Amer. math. Soc.*, 48:488–515, 1940.

Alexander A. Gushchin and Uwe Küchler. Asymptotic inference for a linear stochastic differential equation with time delay. *Bernoulli*, 5(6):1059–1098, 1999.

Alexander A. Gushchin and Uwe Küchler. On stationary solutions of delay differential equations driven by a Lévy process. *Stochastic Processes Appl.*, 88(2):195–211, 2000.

Peter Hall and Christopher C. Heyde. *Martingale limit theory and its application.* Probability and Mathematical Statistics. New York: Academic Press., 1980.

Steven L. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6:327–343, 1993.

Yaozhong Hu and Hongwei Long. Least squares estimator for Ornstein-Uhlenbeck processes driven by α-stable motions. *Stochastic Processes Appl.*, 119(8):2465–2480, 2009.

Jean Jacod. *Calcul stochastique et problèmes de martingales.* Lecture Notes in Mathematics. 714. Berlin-Heidelberg-New York: Springer., 1979.

Jean Jacod and Jean Memin. Caractéristiques locales et conditions de continuite absolue pour les semi-martingales. *Z. Wahrscheinlichkeitstheor. Verw. Geb.*, 35:1–37, 1976.

Jean Jacod and Philip Protter. *Discretization of processes.* Stochastic Modelling and Applied Probability 67. Berlin: Springer. , 2012.

Jean Jacod and Albert N. Shiryaev. *Limit theorems for stochastic processes. 2nd ed.* Grundlehren der Mathematischen Wissenschaften. 288. Berlin: Springer, 2003.

Patrick Jahn, Jørn Hounsgaard Rune W. Berg, and Susanne Ditlevsen. Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of Computational Neuroscience*, 31:563–579, 2011.

G. Jongbloed, F.H. van der Meulen, and A.W. van der Vaart. Nonparametric inference for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 11(5):759–791, 2005.

Geurt Jongbloed and Frank H. van der Meulen. Parametric estimation for subordinators and induced OU processes. *Scand. J. Stat.*, 33(4):825–847, 2006.

Jan Kallsen. A didactic note on affine stochastic volatility models. Kabanov, Yuri (ed.) et al., From stochastic calculus to mathematical finance. The Shiryaev Festschrift., 2006.

Johanna Kappus. *Nonparametric adaptive estimation for discretely observed Lévy processes.* PhD thesis, Humboldt-Universität zu Berlin, 2012.

Achim Klenke. *Probability theory. A comprehensive course.* Universitext. London: Springer., 2008.

A. Kolmogoroff. Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung. *Math. Ann.*, 104:415–458, 1931.

Uwe Küchler and Michael Sørensen. Exponential families of stochastic processes: A unifying semimartingale approach. *Int. Stat. Rev.*, 57(2):123–144, 1989.

Uwe Küchler and Michael Sørensen. *Exponential families of stochastic processes.* Springer Series in Statistics. New York: Springer., 1997.

Uwe Küchler and Michael Sørensen. A note on limit theorems for multivariate martingales. *Bernoulli*, 5(3):483–493, 1999.

Uwe Küchler and Michael Sørensen. A simple estimator for discrete-time samples from affine stochastic delay differential equations. *Stat. Inference Stoch. Process.*, 13(2):125–132, 2010.

Hiroshi Kunita. Itô's stochastic calculus: its surprising power for applications. *Stochastic Processes Appl.*, 120(5):622–652, 2010.

Yury A. Kutoyants. *Statistical inference for ergodic diffusion processes.* Springer Series in Statistics. London: Springer., 2004.

Petr Lansky and Vera Lanska. Diffusion approximations of the neuronal model with synaptic reversal potentials. *Biological Cybernetics*, 56:19–26, 1987.

Lucien Le Cam. Locally asymptotically normal families of distributions. Certain approximations to families of distributions and their use in the theory of estimation and testing hypotheses. *Univ. California Publ. Stat.*, 3:37–98, 1960.

Robert S. Liptser and Albert N. Shiryaev. *Statistics of random processes. 1: General theory. 2nd ed.* Applications of Mathematics. 5. Berlin: Springer., 2001.

Robert Sh. Liptser. A strong law of large numbers for local martingales. *Stochastics*, 3:217–228, 1980.

Michel Loeve. *Probability theory I. 4th ed.* Graduate Texts in Mathematics. 45. New York - Heidelberg - Berlin: Springer., 1977.

Hilmar Mai. *Drift estimation for jump diffusions: time-continuous and high-frequency observations.* PhD thesis, Humboldt-Universität zu Berlin, 2012.

Cecilia Mancini. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scand. J. Stat.*, 36(2):270–296, 2009.

Cecilia Mancini. The speed of convergence of the threshold estimator of integrated variance. *Stochastic Processes Appl.*, 121(4):845–855, 2011.

Hiroki Masuda. On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, 10(1):97–120, 2004.

Hiroki Masuda. Ergodicity and exponential β-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes Appl.*, 117(1):35–56, 2007.

Hiroki Masuda. Approximate self-weighted lad estimation of discretely observed ergodic ornstein-uhlenbeck processes. *Electron. J. Stat.*, 4:525–565, 2010.

Eberhard Mayerhofer, Oliver Pfaffel, and Robert Stelzer. On strong solutions for positive definite jump-diffusions. *Stoch. Proc. Appl.*, 121(9):2072–2086, 2011.

Robert Merton. Option pricing when underlying stock returns are discontinu- ous. *J. Financial Economics*, 3:125–144, 1976.

Salah-Eldin A. Mohammed and Michael K.R. Scheutzow. Lyapunov exponents and stationary solutions for affine stochastic delay equations. *Stochastics Stochastics Rep.*, 29(2):259–283, 1990.

Alexander A. Novikov. Martingales and first-passage times for Ornstein-Uhlenbeck processes with a jump component. *Theory Probab. Appl.*, 48(2):288–303, 2003.

Philip E. Protter. *Stochastic integration and differential equations. 2nd ed.* Applications of Mathematics 21. Berlin: Springer, 2004.

Markus Reiß. Asymptotic equivalence for inference on the volatility from noisy observations. *Ann. Stat.*, 39(2):772–802, 2011.

Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion. 3rd ed., 3rd. corrected printing.* Grundlehren der Mathematischen Wissenschaften 293. Berlin: Springer., 2005.

Ken-Iti Sato. *Lévy processes and infinitely divisible distributions.* Cambridge Studies in Advanced Mathematics. 68. Cambridge: Cambridge University Press., 1999.

Ken-iti Sato and Makoto Yamazato. Operator-selfdecomposable distributions as limit distributions of processes of Ornstein-Uhlenbeck type. *Stochastic Processes Appl.*, 17:73–100, 1984.

Tokuzo Shiga. A recurrence criterion for Markov processes of Ornstein-Uhlenbeck type. *Probab. Theory Relat. Fields*, 85(4):425–447, 1990.

Albert N. Shiryaev. *Probability. 2nd ed.* Graduate Texts in Mathematics. 95. New York: Springer. , 1995.

Albert N. Shiryaev and V.G. Spokoiny. *Statistical experiments and decisions. Asymptotic theory.* Advanced Series on Statistical Science & Applied Probability. 8. Singapore: World Scientific., 2000.

Michael Sørensen. Likelihood methods for diffusions with jumps. *Statistical inference in stochastic processes*, 1991.

Daniel W. Stroock and S. R. Srinivasa Varadhan. *Multidimensional diffusion processes. Reprint of the 2nd correted printing (1997).* Classics in Mathematics. Berlin: Springer., 2006.

Aad W. Van der Vaart. *Asymptotic statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press., 1998.

Abraham Wald. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Trans. Am. Math. Soc.*, 54(3):426–482, 1943.

List of Figures

List of Tables

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Berlin, den 24.04.2012 Hilmar Mai