Technical University of Denmark

Efficient Estimating Functions for Stochastic Differential Equations

Jakobsen, Nina Munkholt

Publication date:
2015

Document Version
Publisher's PDF, also known as Version of record

Efficient Estimating Functions for Stochastic Differential Equations

Jakobsen, Nina Munkholt

Efficient Estimating Functions for Stochastic Differential Equations

Nina Munkholt Jakobsen

This thesis has been submitted to the PhD School of
the Faculty of Science, University of Copenhagen

Nina Munkholt Jakobsen

Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark

munkholt@math.ku.dk // nina@jakobsen.vg

This PhD thesis was submitted to the PhD School of The Faculty of Science, University of Copenhagen, Denmark on 31 March 2015.

Academic advisor:	Michael Sørensen Professor, University of Copenhagen, Denmark

Assessment Committee:	Susanne Ditlevsen (chairman) Professor, University of Copenhagen, Denmark
	Arnaud Gloter Professor, Université d'Évry Val d'Essonne, France
	Hiroki Masuda Associate Professor, Kyushu University, Japan

ISBN: 978-87-7078-960-8

Abstract

The overall topic of this thesis is approximate martingale estimating function-based estimation for solutions of stochastic differential equations, sampled at high frequency. Focus lies on the asymptotic properties of the estimators. The first part of the thesis deals with diffusions observed over a fixed time interval. Rate optimal and efficient estimators are obtained for a one-dimensional diffusion parameter. Stable convergence in distribution is used to achieve a practically applicable Gaussian limit distribution for suitably normalised estimators. In a simulation example, the limit distributions of an efficient and an inefficient estimator are compared graphically. The second part of the thesis concerns diffusions with finite-activity jumps, observed over an increasing interval with terminal sampling time going to infinity. Asymptotic distribution results are derived for consistent estimators of a general multidimensional parameter. Conditions for rate optimality and efficiency of estimators of drift-jump and diffusion parameters are given in some special cases. These conditions are found to extend the pre-existing conditions applicable to continuous diffusions, and impose much stronger requirements on the estimating functions in the presence of jumps. Certain implications of these conditions are discussed, as is a heuristic notion of how efficient estimating functions might be constructed, thus setting the stage for further research.

Preface

This PhD thesis contains the following two papers, which may be read independently:

Jakobsen, N. M. and Sørensen, M. (2015a). Efficient estimation for diffusions sampled at high frequency over a fixed time interval. Preprint.

Jakobsen, N. M. and Sørensen, M. (2015b). Efficient estimation for diffusions with jumps sampled at high frequency over an increasing time interval. Preprint.

The research presented in these papers was done in collaboration with my supervisor Michael Sørensen, Professor at the University of Copenhagen. Both papers are intended for journal publication and were, in their present versions, written mainly by myself, with comments and input from Michael Sørensen.

Acknowledgements

The path to a PhD leads through a rugged mountainous terrain. The journey is scenic and exhilarating, but not without significant challenges along the way. I would like to extend a profound thanks to the following people for assisting me in this endeavour:

Michael Sørensen, for encouraging me to embark on this journey within the field of theoretical statistics. Especially his skilled guidance, thorough feedback and never ending support deserve praise. Our countless discussions were always inspirational and productive. During the most challenging parts of the research project, his energy, vision and positive attitude kept me in high spirits.

Martin Jacobsen, Professor Emeritus at the University of Copenhagen, for contributing valuable thoughts and perspectives during a crucial stage of my research, thus helping me gain insight into the intuition behind some seemingly enigmatic results.

Professor Dr. Markus Reiss, at the Humboldt-Universität zu Berlin, and Dr. Almut Veraart, Reader at Imperial College London, for their kind hospitality during my research stays at their respective departments. Both went above and beyond in sharing their valuable time with me, inspiring me to go in new directions, and broadening my perspectives on a variety of topics.

Special thanks go to my colleagues at the Department of Mathematical Sciences, for creating an inspirational, lively and efficient research environment. Each of them enriched and facilitated my journey in their individual way. Their generosity, as well as their continuous support and trust in my abilities means a lot to me.

My heartfelt thanks go also to my family and friends for always being there for me, for believing in me, and for helping me in any way possible. I am deeply grateful for their patience and understanding. Especially at those crucial times, when I had to devote myself to work. In particular, I thank my husband, Tim, for his loving support, and for making my world perfect.

Summary

Diffusions with and without jumps find wide use in the modelling of dynamical phenomena in continuous time, thus creating a demand for statistical methods to analyse the accompanying data. Although the models have continuous-time dynamics, data can usually only be sampled in discrete time. This complicates the statistical analysis, especially in the presence of jumps. Except in some simple cases, the likelihood function is not known explicitly. Thus, maximum likelihood estimation is generally rendered somewhat impracticable.

The overall topic of this thesis is parametric estimation for stochastic differential equation models with and without jumps, which is carried out using approximate martingale estimating functions. More specifically, focus lies on asymptotic theory for the estimators, which are desired to be rate optimal and efficient.

This thesis essentially consists of two parts. The first part deals with univariate diffusions (without jumps), observed at high frequency over a fixed time interval. These processes are assumed to solve stochastic differential equations with an unknown one-dimensional parameter present in the diffusion coefficient. Existence, uniqueness and asymptotic distribution results are derived for the estimators. The estimators are found to be rate optimal and, under a simple, additional condition, efficient in a local asymptotic mixed normality sense. Stable convergence in distribution is used to obtain a practically applicable standard Gaussian limit distribution for suitably normalised estimators. A concrete example of an efficient approximate martingale estimating function is given, and it is argued that others may be found in the literature. Finally, a small simulation study is used to exemplify the theory, and to compare an efficient and an inefficient estimator graphically.

The second part of the thesis concerns diffusions with finite-activity jumps. These processes are assumed to be observed at high frequency over an increasing time interval, with terminal sampling time going to infinity as the sample size goes to infinity. These processes are also given as solutions to stochastic differential equations, initially, with a general multidimensional parameter allowed to be present in the drift, diffusion and jump coefficients. Again, existence, uniqueness and asymptotic distribution results are obtained for the estimators. Rate optimality and efficiency criteria are motivated by various results in the literature. Subsequently, conditions are given for rate optimality and efficiency of the estimators in three classes of sub-models with unknown drift-jump parameters and/or diffusion parameters. These conditions are found to extend the pre-existing conditions applicable to continuous diffusions, and they impose considerably stronger requirements on the estimating functions in the presence of jumps. Certain implications of these conditions are discussed, as is a heuristic notion of how efficient estimating functions could be constructed, thus setting the stage for further research.

Resumé

Diffusioner med og uden spring finder bred anvendelse i modelleringen af dynamiske fænomener i kontinuert tid. Der skabes derved en efterspørgsel efter statistiske metoder, som kan bruges til at analysere de tilhørende data. Anvendelserne har det tilfælles, at mens modellerne beskriver en udvikling i kontinuert tid, så kan data typisk kun observeres i diskret tid. Dette besværliggør den statistiske analyse, specielt når modellen inkluderer spring. Almindeligvis kendes likelihoodfunktionen for de diskrete observationer ikke eksplicit, hvilket gør maksimum likelihood estimation uanvendeligt i praksis.

Afhandlingens overordnede emne er parametrisk estimation for stokastiske differentialligningsmodeller med og uden spring, som udføres ved brug af approksimative martingal estimationsfunktioner. Der fokuseres på asymptotisk teori for estimatorerne, som specielt ønskes at være rateoptimale og efficiente.

Ud over det indledende kapitel består afhandlingen af to hoveddele. Første del omhandler endimensionelle diffusioner (uden spring), som er observeret ved høj frekvens over et fast tidsinterval. Processerne antages at være løsninger til stokastiske differentialligninger, i hvilke der indgår en ukendt parameter i diffusionskoefficienten. Der etableres eksistens- og entydighedsresultater, samt asymptotiske fordelingsresultater for estimatorerne. Det ses at estimatorerne er rateoptimale, og under yderligere én betingelse er de også efficiente i en lokal asymptotisk normalitets-forstand. Stabil konvergens i fordeling benyttes med henblik på at opnå en praktisk anvendelig grænsefordeling for passende transformerede estimatorer. Ét eksempel gives på en efficient approksimativ martingal estimationsfunktion, og der bliver argumenteret for, at flere eksempler findes i den statistiske litteratur. Til sidst præsenteres et simulationsbaseret eksempel på teorien, hvori der laves grafiske sammenligninger af de asymptotiske fordelinger hørende til henholdsvis en efficient og en inefficient estimator.

Anden del af afhandlingen handler om diffusioner med spring, som ligeledes er observeret med høj frekvens, men over et interval hvor sluttidspunktet for observationerne går mod uendelig, når antallet af observationer går mod uendelig. Disse processer antages også at løse stokastiske differentialligninger, som udgangspunkt med en flerdimensionel parameter som må være til stede både i drifts-, diffusions- og spring-koefficienterne. Igen etableres der eksistens-, entydigheds- og asymptotiske fordelingsresultater for estimatorerne. Kriterier for rateoptimalitet motiveres ud fra forskellige resultater i litteraturen for diffusioner med og uden spring. Der fremsættes betingelser for rateoptimalitet og efficiens i tre typer delmodeller med ukendt drift-spring- og/eller diffusionsparameter. Betingelserne udvider de allerede eksisterende tilsvarende betingelser for kontinuerte diffusioner, men stiller noget højere krav til estimationsfunktionerne, når der også er spring i modellen. Nogle konsekvenser af disse betingelser diskuteres, sammen med en idé til hvordan konkrete eksempler på efficiente approksimative martingal estimationsfunktioner potentielt kunne konstrueres. Samlet lægger disse betragtninger op til videre forskning på området.

Contents

CHAPTER 1

Overview

The overall topic of this thesis is parametric estimation for stochastic differential equations, by means of approximate martingale estimating functions. Focus lies on asymptotic theory for the estimators, which are desired to be rate optimal and efficient. Two further chapters follow: one concerning diffusions without jumps (Chapter 2), the other concerning diffusions with finite-activity jumps (Chapter 3). These chapters correspond to Jakobsen and Sørensen (2015a,b). Each may be read separately, although it should be noted that the bibliography is collected at the end of the thesis.

The rest of this chapter is organised as follows. It commences with a brief background on estimation for diffusions with and without jumps in Section 1.1. Sections 1.2 and 1.3 provide overviews of Chapters 2 and 3, respectively. Each of these sections is divided into three parts. The first part clarifies the objectives of the chapter in question, while the second part summarises the main results achieved. Both parts are set in the context of related literature. The third part serves as a conclusion of the chapter, with perspectives for further research.

1.1 Introduction

Diffusions with and without jumps find wide use in the modelling of dynamical phenomena in continuous time, thus creating a demand for statistical methods to analyse the accompanying data. Some examples of fields of application are agronomy (Pedersen, 2000), biology (Favetto and Samson, 2010), finance (Cox et al., 1985; De Jong et al., 2001; Kou, 2002; Merton, 1971, 1976; Vasicek, 1977) and neuroscience (Bibbona et al., 2010; Ditlevsen and Lansky, 2006; Giraudo and Sacerdote, 1997; Jahn et al., 2011; Musila and Lánský, 1991; Patel and Kosko, 2008; Picchini et al., 2008).

A shared feature of these applications is that although the models have continuous-time dynamics, data can usually only be sampled in discrete time. This complicates the statistical analysis, especially in the presence of jumps.

A *diffusion*, or, in case of ambiguity, a *continuous diffusion* or a *diffusion without jumps*, is defined as the solution $\mathbf{X} = (X_t)_{t\geq 0}$ to a stochastic differential equation of the form

$$dX_t = a(X_t;\theta)\,dt + b(X_t;\theta)\,dW_t\,, \tag{1.1.1}$$

where $\mathbf{W} = (W_t)_{t\geq 0}$ is a standard Wiener process. In the usual parametric setting, the *drift* and *diffusion coefficients*, a and b respectively, are known, deterministic functions of (y,θ), where θ is the unknown, finite-dimensional parameter to be estimated.

Let $(X_{t_0^n}, X_{t_1^n}, \dots, X_{t_n^n})$ denote $n+1$ discrete-time observations of $\mathbf{X}$ at times $0 = t_0^n < t_1^n < \cdots < t_n^n$. Under appropriate assumptions, Markov properties of $\mathbf{X}$ may be used to write the corresponding log-likelihood function, conditional on $X_{t_0^n}$, as

$$\ell_n(\theta) = \sum_{i=1}^{n} \log p(t_i^n - t_{i-1}^n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \tag{1.1.2}$$

with score function

$$\partial_\theta \ell_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(t_i^n - t_{i-1}^n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)\,. \tag{1.1.3}$$

The function $y \mapsto p(\Delta, y, x; \theta)$ represents the transition density, i.e. the conditional density of $X_{t+\Delta}$ given $X_t = x$. However, except in some simple cases, these transition densities are not known explicitly, rendering maximum likelihood estimation somewhat impracticable in general.

A large number of alternate parametric estimation procedures based on discrete observations have been suggested in the literature, many of which perform well under various sampling scenarios. A non-exhaustive list of references is presented in the following. For further reference, see also the overview given by Sørensen (2004).

Pseudo-likelihood methods, i.e. approximations of the likelihood or log-likelihood functions, often of a Gaussian type, were considered by, e.g. Florens-Zmirou (1989), Genon-Catalot (1990), Genon-Catalot and Jacod (1993), Gloter and Sørensen (2009), Jacod (2006), Kessler (1997), Prakasa Rao (1983), Sørensen and Uchida (2003), and Yoshida (1992). Aït-Sahalia (2002, 2008), Dacunha-Castelle and Florens-Zmirou (1986), and Li (2013) focused on expansions of the transition densities, while the approaches of, e.g. Bibby and Sørensen (1995), Jacobsen (2001, 2002), Sørensen (2010), and Uchida (2004, 2008) concerned approximation of the score function. Furthermore, simulation-based likelihood methods were considered by, e.g. Beskos et al. (2006, 2009), Durham and Gallant (2002), Pedersen (1995) and Roberts and Stramer (2001).

There also exist a number of non-parametric estimation procedures based on discrete observations. That is, methods designed for diffusion models where, e.g. the drift and diffusion coefficients a and b themselves are unknown functions to be estimated. For references, see e.g. Bandi and Phillips (2003), Comte et al. (2007), Florens-Zmirou (1993), Genon-Catalot et al. (1992), Jacod (2000) and Schmisser (2013). Recently, the development of Bayesian non-parametric methods was the focus of, e.g. Papaspiliopoulos et al. (2012), van der Meulen and van Zanten (2013), and van der Meulen et al. (2014).

A *diffusion with jumps* is defined as the (càdlàg) solution $\mathbf{X} = (X_t)_{t\geq 0}$ to the stochastic differential equation

$$dX_t = \tilde{a}(X_t; \theta)\, dt + b(X_t; \theta)\, dW_t + \int_{\mathbb{R}} c(X_{t-}, z; \theta)\, (N^\theta - \mu_\theta)(dt, dz)\,, \tag{1.1.4}$$

a generalisation of (1.1.1), where $\mathbf{X}_- = (X_{t-})_{t\geq 0}$ denotes the process of left limits. The time-homogeneous Poisson random measure $N^\theta(dt, dz)$ is independent of $\mathbf{W}$, and has the intensity measure $\mu_\theta(dt, dz) = \nu_\theta(dz)\,dt$ for some Lévy measure ν_θ. In extension to the description surrounding (1.1.1), in the fully parametric framework, the measure ν_θ is usually known up to the parameter θ. The *jump coefficient* c is a known, deterministic function of $(y, z; \theta)$. The *compensated drift coefficient* $\tilde{a}$ may be written as

$$\tilde{a}(y;\theta) = a(y;\theta) + \int_{\mathbb{R}} c(y, z; \theta)\,\nu_\theta(dz)\,,$$

when the integral exists. When $\nu_\theta(\mathbb{R}) < \infty$, the jumps of $\mathbf{X}$ are said to be of *finite activity*. In this case, (1.1.4) may also be represented as

$$dX_t = a(X_t;\theta)\,dt + b(X_t;\theta)\,dW_t + \int_{\mathbb{R}} c(X_{t-}, z; \theta)\,N^\theta(dt, dz)\,, \tag{1.1.5}$$

and $\mathbf{X}$ is often referred to as a *jump-diffusion*. As opposed to a diffusion, which has continuous sample paths (with probability one), a jump-diffusion exhibits at most finitely many jumps in any time interval of finite length. In intervals without jumps, it follows the dynamics given by (1.1.1). When $\nu_\theta(\mathbb{R}) = \infty$, the jumps of $\mathbf{X}$ are said to be of *infinite activity*, in which case $\mathbf{X}$ jumps infinitely many times in any finite time interval.

Under appropriate conditions, a diffusion with jumps is also a Markov process, and the expressions (1.1.2) and (1.1.3) are still valid. However, the challenge of finding an analytic expression for the transition density is no less great than for continuous diffusions. In the absence of a closed-form expression for the log-likelihood function, statistical inference is complicated further by the following: To the extent that knowledge of the jump times and sizes is needed, it has to be inferred from the discrete-time observations whether one or more jumps are likely to have occurred between any two consecutive observation times, and, if so, how much of the observed increment is attributable to the jump(s). This information would be more easily obtainable from the ideal continuous-time observations, at least, in the case of finite-activity jumps.

Again, a multitude of estimation approaches may be found in the literature. A non-exhaustive list of references includes the following: In the context of parametric estimation, pseudo-likelihood methods, primarily involving Gaussian approximations to the log-likelihood (or score) function, were considered by, e.g. Masuda (2011, 2013), Ogihara and Yoshida (2011), Shimizu (2006b), and Shimizu and Yoshida (2006). Closed-form expansion of the transition densities was investigated by, e.g. Filipović et al. (2013), and Yu (2007), while Mai (2014) approximated maximum likelihood estimators obtained from the continuous-time likelihood function. Mancini (2004) proposed a quadratic variation-inspired estimation method in a semiparametric setting, while simulation-based methods were considered by, e.g. Giesecke and Schwenkler (2014), and Stramer et al. (2010). Finally, a selection of non-parametric procedures based on discrete observations exist as well, see e.g. Bandi and Nguyen (2003), Mancini (2009), Mancini and Renò (2011), Schmisser (2014) and Shimizu (2006a, 2008, 2009).

1.2 Diffusions Without Jumps

1.2.1 Background and Objectives

In Chapter 2, we consider continuous diffusions **X** solving stochastic differential equations of the form

$$dX_t = a(X_t)\,dt + b(X_t;\theta)\,dW_t \tag{1.2.1}$$

for $\theta \in \Theta$. These constitute a special case of (1.1.1), where the unknown parameter is only present in the diffusion coefficient. In the following, the true, unknown parameter is denoted θ_0. For $n \in \mathbb{N}$, it is assumed that **X** is observed at $n+1$ discrete, equidistant time-points $t_i^n = i/n$, $i = 0, 1, \ldots, n$, over the fixed interval $[0, 1]$. In the following, asymptotics are considered as $n \to \infty$. We say that **X** is observed at *high frequency*, because the time-distance $\Delta_n = t_i^n - t_{i-1}^n$ satisfies that $\Delta_n = 1/n \to 0$ as $n \to \infty$.

For simplicity, X_t and θ are both assumed to be one-dimensional. Extension of our results to a multivariate parameter is expected to be quite straightforward. Drift parameters cannot be estimated consistently under the fixed-interval asymptotic scenario considered here, and are therefore excluded from the model. The choice of time-interval $[0, 1]$ is not considered restrictive, as the results may be generalised to other compact intervals by suitable rescaling of the drift and diffusion coefficients.

In this setup, the *local asymptotic mixed normality* (LAMN) property has been shown to hold (Dohnal, 1987; Gobet, 2001)[1] with rate $\sqrt{n}$ and random asymptotic Fisher information

$$\mathcal{I}(\theta_0) = 2\int_0^1 \left(\frac{\partial_\theta b(X_s;\theta_0)}{b(X_s;\theta_0)}\right)^2 ds = \frac{1}{2}\int_0^1 \left(\frac{\partial_\theta b^2(X_s;\theta_0)}{b^2(X_s;\theta_0)}\right)^2 ds\,. \tag{1.2.2}$$

Here, e.g. $\partial_\theta b^2(x;\theta)$ denotes the partial derivative of b^2 with respect to θ. In the context of local asymptotic mixed normality, a consistent estimator $\hat{\theta}_n$ of θ_0 is *rate optimal* if $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to a non-degenerate random variable. Furthermore, it is *efficient* if this limit distribution may be written on the form $\mathcal{I}(\theta_0)^{-1/2}Z$, where Z follows a standard normal distribution, and is independent of $\mathcal{I}(\theta_0)$. In general terms, over all consistent estimators $\hat{\theta}_n$, the optimal rate of convergence $\delta_n = \sqrt{n}$ is the "fastest possible" rate at which $\delta_n(\hat{\theta}_n - \theta_0)$ converges in distribution to a non-degenerate limit. Similarly, conditional on $\mathcal{I}(\theta_0)$, the distribution characterised by $\mathcal{I}(\theta_0)^{-1/2}Z$ has the "smallest conditional variance possible", for a limit distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$. (See Section 2.2.6 of Chapter 2 for further details.)

Much of the literature on parametric estimation for diffusions concerns sampling scenarios where $n\Delta_n \to \infty$ as $n \to \infty$, with either $\Delta_n \to 0$ (high frequency asymptotics) or $\Delta_n = \Delta$ fixed (low frequency asymptotics). In these cases, drift and diffusion parameters can both be estimated consistently. Limit distributions of suitably normalised estimators are generally Gaussian, with variances that depend on the true unknown parameter θ_0,

[1] Dohnal considered univariate diffusions, Gobet multivariate diffusions.

and on Δ in the case of low frequency asymptotics. See, e.g. the asymptotic results of Dacunha-Castelle and Florens-Zmirou (1986), Florens-Zmirou (1989), Jacobsen (2001), Kessler (1997), Sørensen (2010), and Yoshida (1992).

Gaussian limit distributions are also obtained within the framework of small-diffusion asymptotics, as studied in the papers of, e.g. Genon-Catalot (1990), Gloter and Sørensen (2009), Sørensen and Uchida (2003), and Uchida (2004, 2008). Small-diffusion asymptotics entail $\Delta_n \to 0$ with $n\Delta_n$ fixed, as in the current setting, but under the additional assumption that the diffusion coefficient is of the form $b(y;\theta) = \varepsilon\tilde{b}(y;\theta)$, with $\varepsilon \to 0$ and $n \to \infty$ simultaneously. Unlike in our fixed-interval setting, drift parameters can be estimated consistently, so the drift and/or diffusion coefficient may depend on unknown parameters. The asymptotic variances of suitably normalised estimators generally depend on the path of the corresponding ordinary differential equation under the true parameter, obtained by setting $\varepsilon = 0$.

In the current setting, where the asymptotics consist of $\Delta_n = 1/n \to 0$ as $n \to \infty$ with $n\Delta_n$ fixed, the limit distributions of consistent estimators tend to be more complicated. This is not only because they are generally seen to be normal variance mixtures. Even just for efficient estimators, it is seen from (1.2.2) that the distributions typically depend on $(X_t)_{t\in[0,1]}$, the full sample path of the the diffusion process over the observation interval, which is only partially observed in practice. Parametric estimation under this particular asymptotic scenario has previously been considered by Genon-Catalot and Jacod (1993, 1994) and, to some extent, by Dohnal (1987) and Jacod (2006), in addition to the local asymptotic mixed normality results of Dohnal (1987) and Gobet (2001).

The setup described here is a special case of the one considered by Genon-Catalot and Jacod (1993), who proposed estimators of the diffusion parameter based on a class of contrast functions.[2] When adapted to our framework, these contrast functions have the form

$$U_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} f\left(b^2(X_{t_{i-1}^n};\theta), \Delta_n^{-1/2}(X_{t_i^n} - X_{t_{i-1}^n})\right),$$

for functions $f(v,w)$ satisfying certain conditions, and may thus only depend on the observations through $b^2(X_{t_{i-1}^n};\theta)$ and $\Delta_n^{-1/2}(X_{t_i^n} - X_{t_{i-1}^n})$. Estimators of θ_0 are obtained by minimising the contrast functions. They are seen to be rate optimal, and, when suitably normalised, they converge in distribution to normal variance mixtures, which generally depend on the sample path $(X_t)_{t\in[0,1]}$ (Genon-Catalot and Jacod, 1993, Theorem 3). The contrast function based on $f(v,w) = \log v + w^2/v$ was identified as efficient (Genon-Catalot and Jacod, 1993, Theorem 5).

After showing the local asymptotic mixed normality property for the current model and observation scheme, Dohnal noted that when $b^2(x;\theta) = h(x)k(\theta)$ for appropriate functions h

[2]The paper of Genon-Catalot and Jacod (1994) generalised their results from 1993, in the sense that the paper from 1994 focused on random sampling times (and is thus out of the scope of this thesis). In extension of what is included in the following discussion, their paper from 1993 also allowed non-equidistant sampling times, multi-dimensional diffusion parameters and multivariate processes with more general drift coefficients than ours.

and k, the efficient limit distribution reduces to a normal distribution, and the local asymptotic mixed normality to local asymptotic normality (LAN). (See also (Genon-Catalot and Jacod, 1993, Example 7.b).) Indeed, it is seen from (1.2.2) that $\mathcal{I}(\theta_0)^{-1} = 2k^2(\theta_0)/\partial_\theta k(\theta_0)^2$ becomes the (non-random) covariance matrix of the asymptotic distribution in this case. For example, the squared diffusion coefficients of a number of Pearson diffusions, such as the Ornstein-Uhlenbeck and square root processes, may be written on the specified product form. (See Forman and Sørensen (2008) for more on Pearson diffusions.) Dohnal considered two examples of sub-models of (1.2.1) for which the squared diffusion coefficient was such a product (one of them an Ornstein-Uhlenbeck process), and proposed some explicit, efficient estimators for these models, based on their local asymptotic normality.

Furthermore, as part of a more general paper on asymptotics for estimators of parameters in non-ergodic diffusions, Jacod (2006, Theorem 2.1) proposed a contrast function for estimating the diffusion parameter within the present setting. He argued that for the resulting estimators $\hat{\theta}_n$, $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is tight.

On a similar note, in the non-parametric literature, normal variance mixture limit distributions have been observed as well in connection with related estimation problems. For example, when estimating integrated volatility $\int_0^1 b^2(X_s)\,ds$ (Jacod and Protter, 1998; Mykland and Zhang, 2006), or the squared diffusion coefficient $b^2(x)$ over a fixed interval (Florens-Zmirou, 1993; Jacod, 2000).

Our main objective in Chapter 2 is to establish the existence of rate optimal and efficient estimators of θ_0 within the general model (1.2.1), based on the extensive class of approximate martingale estimating functions. We also aim to find a suitable (data-dependent) normalisation of these estimators, which converges in distribution to a practically applicable limit distribution, that does not depend on any unknown or unobserved quantities.

Approximate martingale estimating functions, defined more precisely in Sections 2.2.3 and 3.2.3, may be written as e.g.

$$G_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta),$$

where $g(t, y, x; \theta)$ is a deterministic function. Here $g(t, y, x; \theta)$ is real-valued, whereas for a d-dimensional parameter θ, it would be $\mathbb{R}^d$-valued. The approximate martingale property assumed to be satisfied by g is a conditional expectation condition of the form

$$\mathbb{E}_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}) = \Delta_n^\kappa R_\theta(\Delta_n, X_{t_{i-1}^n}) \tag{1.2.3}$$

for some $\kappa \geq 2$, where the remainder term $R_\theta(t, x)$ may be controlled as necessary. Estimators based on $G_n(\theta)$ are referred to as G_n-estimators, and are essentially obtained as solutions to the estimating equation $G_n(\theta) = 0$. Estimating functions of this type were used by e.g. Bibby and Sørensen (1995), Jacobsen (2001, 2002), Sørensen (2010), and Uchida (2004), in connection with other diffusion models and asymptotic schemes.

The model (1.2.1) is a sub-model of that studied by Sørensen (2010), which included a drift parameter. Sørensen considered estimation by approximate martingale estimating functions under the asymptotic scenario $\Delta_n \to 0$ and $t_n^n = n\Delta_n \to \infty$ as $n \to \infty$. Not only did he give simple conditions for rate optimality and efficiency (in the local asymptotic normality sense), he also argued that the theory of approximate martingale estimating functions encompasses a large number of other well-performing estimators in the literature.

Based on the results of Sørensen (2010), it is important, in our opinion, to investigate the performance of estimators based on approximate martingale estimating functions under the present high-frequency, fixed-interval observation scheme. We hope to find similar, simple conditions for rate optimality and efficiency as those found by Sørensen.

1.2.2 Overview of Main Results

In the following, convergence in distribution and in probability, denoted $\xrightarrow{\mathcal{D}}$ and $\xrightarrow{\mathcal{P}}$ respectively, are understood to be under the true probability measure as $n \to \infty$. Furthermore, for example, $\partial_y^2 g(0, x, x; \theta)$ denotes the second partial derivative of $g(0, y, x; \theta)$ with respect to y, evaluated in $y = x$.

The first main contribution of Chapter 2 is Theorem 2.3.2, which establishes existence, uniqueness, and asymptotic distribution results for rate optimal G_n-estimators, within the setup described in Section 1.2.1. It also shows that suitably normalised estimators converge in distribution to a standard Gaussian limit suitable for practical purposes, in that knowledge of the full sample path $(X_t)_{t\in[0,1]}$ is no longer needed. Omitting the technical details and regularity assumptions, the theorem may be summarised as follows:

Theorem. *Suppose that the appropriate assumptions hold. Then,*

(i) there exists a consistent G_n-estimator $\hat{\theta}_n$. In any compact, convex set $K \subseteq \Theta$ containing θ_0 in its interior, the estimator is unique with probability going to one.

(ii) for any consistent G_n-estimator $\hat{\theta}_n$, it holds that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} W(\theta_0)Z\,. \tag{1.2.4}$$

Z follows a standard normal distribution and is independent of $W(\theta_0)$, given by

$$W(\theta_0) = \frac{\left(\int_0^1 \frac{1}{2} b^4(X_s;\theta_0)\partial_y^2 g(0, X_s, X_s;\theta_0)^2\, ds\right)^{1/2}}{\int_0^1 \frac{1}{2}\partial_\theta b^2(X_s;\theta_0)\partial_y^2 g(0, X_s, X_s;\theta_0)\, ds}\,. \tag{1.2.5}$$

A further specified transformation of the observed data, $\widehat{W}_n$, with $\widehat{W}_n \xrightarrow{\mathcal{P}} W(\theta_0)$, satisfies that

$$\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} Z\,. \tag{1.2.6}$$

◇

Due to the general randomness of $W(\theta_0)$, the concept of *stable convergence in distribution* was employed in order to obtain (1.2.6). The rate of convergence in (1.2.4) reveals that the consistent G_n-estimators are rate optimal. This was ensured by imposing the condition

$$\partial_y g(0, x, x; \theta) = 0 \tag{1.2.7}$$

for all x and θ, on the function $g(t, y, x; \theta)$.

The second main contribution in Chapter 2, formally established in Corollary 2.3.4, concerns the efficiency of the G_n-estimators. Using (1.2.5), it is seen that under the additional condition

$$\partial_y^2 g(0, x, x; \theta) = K_\theta \frac{\partial_\theta b^2(x; \theta)}{b^4(x; \theta)} \tag{1.2.8}$$

for all θ and x, where K_θ is a non-zero, possibly θ-dependent constant, any consistent G_n-estimator $\hat{\theta}_n$ is efficient.

As an example of an efficient estimating function, it may easily be verified that the approximate martingale estimating function $\bar{G}_n(\theta)$ given by

$$\bar{g}(t, y, x; \theta) = \frac{\partial_\theta b^2(x; \theta)}{b^4(x; \theta)} \left((y - x)^2 - t b^2(x; \theta) \right)$$

satisfies the conditions (1.2.7) and (1.2.8), and corresponds to the contrast function shown to be efficient by Genon-Catalot and Jacod (1993, Theorem 5). The latter because with $\Delta_n = 1/n$, the efficient contrast function of Genon-Catalot and Jacod may be written on the form $\bar{U}_n(\theta) = \sum_{i=1}^n \bar{u}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$ with

$$\bar{u}(t, y, x; \theta) = t \log b^2(x; \theta) + (y - x)^2 / b^2(x; \theta) \,,$$

and $\partial_\theta \bar{u}(t, y, x; \theta) = -\bar{g}(t, y, x; \theta)$. Thus, $\bar{U}_n(\theta)$ and $\bar{G}_n(\theta)$ yield the same estimators. Furthermore, for the sub-model of (1.2.1) given by $dX_t = a(X_t)\, dt + \sqrt{\theta}\tilde{b}(X_t)\, dW_t$, Dohnal (1987) proposed, e.g. the efficient estimator

$$\tilde{\theta}_n = \sum_{i=1}^n \frac{\left(X_{t_i^n} - X_{t_{i-1}^n} - \Delta_n a(X_{t_{i-1}^n}) \right)^2}{\tilde{b}^2(X_{t_{i-1}^n})} \,.$$

The estimator $\tilde{\theta}_n$ can also be obtained as the unique solution to the estimating equation when using the efficient approximate martingale estimating function given by

$$\tilde{g}(t, y, x; \theta) = \frac{\partial_\theta b^2(x; \theta)}{b^4(x; \theta)} \left((y - x - t a(x))^2 - t b^2(x; \theta) \right) \,.$$

For the model $dX_t = \tilde{a} X_t\, dt + \sqrt{\theta}\, dW_t$ with $\tilde{a}$ known, an efficient estimator proposed by Dohnal was

$$\check{\theta}_n = \frac{1}{4} \sum_{i=1}^n \left((2 + \Delta_n \tilde{a}) X_{t_i^n} - (2 - \Delta_n \tilde{a}) X_{t_{i-1}^n} \right)^2 \,.$$

This estimator is also obtainable as the unique solution to the estimating equation when using an approximate martingale estimating function based on

$$\breve{g}(t, y, x; \theta) = ((2 + t\tilde{a})y - (2 - t\tilde{a})x)^2 - 4t\theta\,,$$

which satisfies the assumptions (1.2.7) and (1.2.8) for the model in question. Lemma 2.2.6 in Chapter 2 may be used to verify that $\bar{g}$, $\tilde{g}$ and $\breve{g}$ satisfy the approximate martingale condition (1.2.3).

The expressions (1.2.7) and (1.2.8) correspond to the conditions found for rate optimality and efficiency of diffusion parameter-estimators within the framework of Sørensen (2010). Furthermore, as discussed by Sørensen, they also emerge in the work of Jacobsen (2002). There, they were given as conditions for small Δ-optimality of martingale estimating functions in the sense of Jacobsen (2001), in models with only a diffusion parameter. Small Δ-optimality concerns the near-efficiency of estimating functions based on discrete observations, with a fixed distance Δ close to 0 between observation times. In general terms, small Δ-optimal estimating functions yield estimators that achieve a lower bound on the asymptotic variance in the limit $\Delta \to 0$. Consequently, a number of (approximate) martingale estimating functions discussed by Jacobsen (2002) and Sørensen (2010) also satisfy our rate optimality and efficiency conditions.

An additional contribution of Chapter 2 is a small simulation study, in which we make graphical comparisons of the distributions of two estimators, one efficient and one not. In accordance with our theoretical considerations, the efficient estimator is seen to have preferable properties. The Gaussian limit distribution in (1.2.6) approximates the distribution of the normalised efficient estimator very well for the sample sizes considered. A more notable discrepancy is seen in the case of the inefficient estimator. It is also illustrated that the limit distribution in (1.2.4) is much more spread out for the inefficient estimator than for the efficient estimator.

1.2.3 Conclusions and Perspectives for Further Research

In Chapter 2, we considered estimation of the diffusion parameter of a continuous diffusion process observed at high frequency over a fixed interval. Existence, uniqueness properties and asymptotic distribution results were established for rate optimal estimators based on approximate martingale estimating functions. Rate optimality was ensured by a simple condition, and a straightforward supplementary condition ensuring efficiency was stated as well. We used stable convergence in distribution to achieve a practically applicable standard Gaussian limit distribution for suitably normalised estimators. An example of an efficient approximate martingale estimating function was given,[3] and it was argued that there exist more approximate martingale estimating functions in the literature, which satisfy our rate optimality and efficiency conditions. Finally, we compared an efficient and an inefficient estimating function by simulation, and saw graphically, that the efficient estimator had preferable properties.

[3] As well as two further examples in Section 1.2.2.

The results obtained were for a univariate diffusion process with a one-dimensional diffusion parameter, observed at equidistant time-points with time step $\Delta_n = 1/n$. Based on our later work presented in Chapter 3 (concerning jump-diffusions with a multidimensional parameter) it should be quite straightforward to extend the results of Chapter 2 to a multi-dimensional diffusion parameter as well. Furthermore, in the paper of Genon-Catalot and Jacod (1993), the stochastic processes considered were also multivariate, and the observation times not necessarily equidistant. Such extensions of our work are likely to be possible as well.

Had time permitted, we would have liked to develop the simulation study further. For example, with applications in mind, it would be useful to compare the finite sample properties of different efficient estimators with each other and with more general rate optimal estimators. More specifically, the following question could be posed: For practically feasible sample sizes, is the Gaussian approximation to the distribution of the normalised estimators generally better for the efficient estimators than for those that are inefficient?

Finally, it would be fascinating to expand the investigation in Chapter 2 to diffusions with jumps as well. In light of our later results in Chapter 3 (for the asymptotic scenario $\Delta_n \to 0$, $n\Delta_n \to \infty$ as $n \to \infty$), this would probably not be so straightforward. However, based on the fine theoretical properties of approximate martingale estimating functions in the case of continuous diffusions, it comes across as an important, yet to our knowledge, unresearched area.

1.3 Diffusions With Jumps

1.3.1 Background and Objectives

In Chapter 3, we consider ergodic diffusions $\mathbf{X}$ with finite-activity jumps ($\nu_\theta(\mathbb{R}) < \infty$) and càdlàg paths, which solve stochastic differential equations of the form

$$dX_t = a(X_t;\theta)\,dt + b(X_t;\theta)\,dW_t + \int_{\mathbb{R}} c(X_{t-}, z;\theta)\,N^\theta(dt, dz) \tag{1.3.1}$$

with $\theta \in \Theta$, as seen in (1.1.5). The invariant distribution of $\mathbf{X}$ is denoted by π_θ, and the true, unknown parameter by θ_0. With $(\Delta_n)_{n\in\mathbb{N}}$ a sequence of strictly positive numbers, it is assumed that for $n \in \mathbb{N}$, $\mathbf{X}$ is observed at $n+1$ discrete, equidistant time-points $t_i^n = i\Delta_n$, $i = 0, 1, \ldots, n$, over the interval $[0, n\Delta_n]$. Asymptotics are considered as $n \to \infty$, in which case it is assumed that $\Delta_n \to 0$ and $n\Delta_n \to \infty$. With this observation scheme, $\mathbf{X}$ is said to be observed at high frequency over an increasing time interval, with terminal sampling time $t_n^n = n\Delta_n$ going to infinity. In the limit $\Delta_n \to 0$, the whole sample path of the process is (hypothetically) observed, containing full information on the jump times and sizes. As in the previous chapter, X_t is assumed to be one-dimensional, whereas θ is assumed to be d-dimensional for some $d \in \mathbb{N}$.

Local asymptotic normality (LAN), and, for fixed-interval asymptotics, local asymptotic mixed normality, are an active area of research for processes with jumps. Recent developments include the work of Becheri et al. (2014), Clément and Gloter (2015), Kawai and

Masuda (2013), and Kohatsu-Higa et al. (2014, 2015). Within the context of local asymptotic normality, it is quite straightforward to characterise rate optimality and efficiency of estimators. However, in the absence of comprehensive local asymptotic normality results for the present setup, the criteria for rate optimality and efficiency used in this paper are more heuristic in nature.

Let $M^\star$ denote transposition of a matrix (or vector) M. Consider the sub-model of (1.3.1) given by

$$dX_t = a(X_t;\alpha)\,dt + b(X_t;\beta)\,dW_t + \int_{\mathbb{R}} c(X_{t-},z;\alpha)\,N^\alpha(dt,dz)\,, \tag{1.3.2}$$

where the unknown parameter θ is split into a drift-jump parameter α and a diffusion parameter β, such that $\theta^\star = (\alpha^\star,\beta^\star)$. For this model, based on results in the literature,[4] we conjecture on the following properties, over all consistent estimators $\hat{\theta}_n$ of θ_0: The "fastest possible" rate of convergence $\delta_n = \delta_{0,n}$ of $\delta_n(\hat{\theta}_n - \theta_0)$ to a non-degenerate limit distribution (for rate optimality), and the smallest possible asymptotic variance of $\delta_{0,n}(\hat{\theta}_n - \theta_0)$ (for efficiency). The matrix δ_n is invertible and diagonal, with diagonal elements satisfying that $(\delta_n)_{jj} \to \infty$ as $n \to \infty$ for all $j = 1, \ldots, d$.

Suppose, for a moment, that the Lévy measure ν_θ has density with respect to Lebesgue measure. Let $w \mapsto \varphi(x,w;\alpha)$ denote the transformation of the Lévy density by $z \mapsto c(x,z;\alpha)$, and put $\mathcal{W}(x) = c(x,\mathbb{R};\alpha)$. In this case, the conjecture (Conjecture 3.4.4) may be summarised as follows:

Conjecture. *Under suitable assumptions, a consistent estimator $\hat{\theta}_n$ of the true, unknown parameter θ_0 is rate optimal if*

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_d(0,\mathcal{V}(\theta_0))\,,$$

and efficient when $\mathcal{V}(\theta_0)$ is the (well-defined) inverse of the block diagonal matrix given by $\mathcal{I}(\theta_0) = blockdiag(\mathcal{I}_1(\theta_0),\mathcal{I}_2(\theta_0))$, with

$$\mathcal{I}_1(\theta_0) = \int_X \left(\frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x;\beta_0)} + \int_{\mathcal{W}(x)} \frac{\partial_\alpha \varphi(x,w;\alpha_0)^\star \partial_\alpha \varphi(x,w;\alpha_0)}{\varphi(x,w;\alpha_0)}\,dw \right) \pi_{\theta_0}(dx)$$

$$\mathcal{I}_2(\theta_0) = \frac{1}{2}\int_X \frac{\partial_\beta b^2(x;\beta_0)^\star \partial_\beta b^2(x;\beta_0)}{b^4(x;\beta_0)}\,\pi_{\theta_0}(dx)\,.$$

◇

Here $\hat{\theta}_n^\star = (\hat{\alpha}_n^\star,\hat{\beta}_n^\star)$, while, e.g. $\partial_\alpha a(x;\alpha)$ denotes the row-vector containing the partial derivatives of $a(x;\alpha)$ with respect to the coordinates of α, and $\mathcal{N}_d(0,V)$ is the d-dimensional, zero-mean Gaussian distribution with covariance matrix V.

[4]In Section 3.4 of Chapter 3, we motivate the Conjecture 3.4.4 using the local asymptotic normality results of Becheri et al. and Kohatsu-Higa et al. (applicable to 1.3.2 in special cases), and other results of Gobet (2002), Shimizu and Yoshida (2006) and Sørensen (1991).

Parametric estimation covering sub-models of (1.3.1) has previously been considered by e.g. Masuda (2011, 2013), Ogihara and Yoshida (2011), Shimizu (2006b), and Shimizu and Yoshida (2006).[5] Shimizu and Yoshida proposed a technique to judge whether or not a jump is likely to have occurred between two observation times t_{i-1}^n and t_i^n. They used this technique to create a contrast function for estimation in sub-models of the form (1.3.2), which may be written as

$$\begin{aligned} H_n(\theta) = &-\frac{1}{2\Delta_n}\sum_{i=1}^n \Big(\Delta X_{n,i} - \Delta_n a(X_{t_{i-1}^n};\alpha)\Big)^2 b^{-2}(X_{t_{i-1}^n};\beta)\mathbf{1}(|\Delta X_{n,i}| \le \Delta_n^\rho) \\ &- \sum_{i=1}^n \frac{1}{2}\Big(\log b^2(X_{t_{i-1}^n};\beta)\Big)\mathbf{1}(|\Delta X_{n,i}| \le \Delta_n^\rho) \\ &+ \sum_{i=1}^n \Big(\log \Phi_n(X_{t_{i-1}^n}, \Delta X_{n,i};\alpha)\Big)\phi_n(X_{t_{i-1}^n}, \Delta X_{n,i})\mathbf{1}(|\Delta X_{n,i}| > \Delta_n^\rho) \\ &- \Delta_n \sum_{i=1}^n \int_{\mathcal{W}_{n,i}} \Phi_n(X_{t_{i-1}^n}, w;\alpha)\,dw\,. \end{aligned} \tag{1.3.3}$$

$\mathbf{1}(A)$ denotes the indicator function of the set A, $\Delta X_{n,i} = X_{t_i^n} - X_{t_{i-1}^n}$, $\phi_n(x,w)$ is a truncation function used to ensure integrability, and $\Phi_n(x,w;\alpha) = \varphi(x,w;\alpha)\phi_n(x,w)$, $\mathcal{W}_{n,i} = \mathcal{W}(X_{t_{i-1}^n})$ with $\varphi(x,w;\alpha)$ and $\mathcal{W}(x)$ as described earlier. For finite sample sizes, the choice of the constant ρ affects the ability of the contrast function to determine whether or not a jump has occurred between t_{i-1}^n and t_i^n.

In their Theorem 2.1, Shimizu and Yoshida established the asymptotic distribution of the estimator obtained by maximising the contrast function $H_n(\theta)$. They argued that the contrast function is efficient for the drift-jump parameter. By the criteria laid out in the aforementioned conjecture, it is also efficient for the diffusion parameter.

Also considering estimation in the model (1.3.2), Ogihara and Yoshida (2011) used a contrast function which was essentially identical to the one of Shimizu and Yoshida. Under weaker assumptions on the Lévy measure, they proved convergence in distribution of the estimator to the efficient limit distribution (their Theorem 1). They also proved convergence of the moments of the estimator to moments of the limit distribution, as well as similar results for a Bayes type estimator based on the same contrast function. Shimizu (2006b) proposed and investigated the asymptotics of an estimating function heavily inspired by the efficient contrast function of Shimizu and Yoshida, but modified with the application to infinite-activity jumps in mind. In general terms, he concluded that his estimator was not efficient for jump parameters.

Masuda (2011, 2013) considered Gaussian quasi-likelihood estimation for diffusions with (possibly infinite-activity) jumps, which, in special cases, overlap with sub-models of (1.3.1) of the form

$$dX_t = a(X_t;\alpha)\,dt + b(X_t;\beta)\,dW_t + \int_{\mathbb{R}} \tilde{c}(X_{t-},\beta)z\,N(dt,dz)\,.$$

[5] Several of these papers assumed multivariate processes and/or allowed infinite-activity jumps. In the following, we mainly refer to their results within the framework of univariate processes with finite-activity jumps.

Among other things, Masuda studied theoretical asymptotics for his Gaussian quasi-likelihood estimators under the current asymptotic scenario. In particular, Theorem 3.4 of Masuda's paper from 2011, and Theorems 2.7 & 2.9 of his paper from 2013 established convergence in distribution of suitably normalised estimators and functions thereof. Masuda pointed out that in the presence of jumps, these estimators are not efficient for the drift or diffusion-jump parameters of the model, or even rate optimal for parameters of the diffusion coefficient. Estimation using Gaussian quasi-likelihood functions of the types considered by Masuda fits into the framework of approximate martingale estimating functions, which were briefly described in Section 1.2.1. For example, the Gaussian quasi-likelihood function of Masuda (2013) corresponds to an estimating function of the form $\bar{G}_n(\theta) = \sum_{i=1}^n \bar{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$ where $\bar{g}^\star = (\bar{g}_\alpha^\star, \bar{g}_\beta^\star)$ and

$$\begin{aligned}\bar{g}_\alpha(t,y,x;\theta) &= \frac{\partial_\alpha a(x;\alpha)^\star}{(b^2+\tilde{c}^2)(x;\beta)}(y-x-ta(x;\alpha))\\ \bar{g}_\beta(t,y,x;\theta) &= \frac{\partial_\beta (b^2+\tilde{c}^2)(x;\beta)^\star}{(b^2+\tilde{c}^2)^2(x;\beta)}\left((y-x-ta(x;\alpha))^2 - t(b^2+\tilde{c}^2)(x;\beta)\right).\end{aligned}$$

Using Lemma 3.2.8 from Chapter 3, it may be verified that under the assumptions of Masuda (2013), $\bar{g}(t,y,x;\theta)$ satisfies the approximate martingale property (1.2.3). However, to our knowledge, the theoretical asymptotic properties of more general approximate martingale estimating functions for diffusions with jumps have not yet been investigated.

The observation scheme considered here matches that of Sørensen (2010). For continuous diffusions of the form

$$dX_t = a(X_t;\alpha)\,dt + b(X_t;\beta)\,dW_t\,,$$

Sørensen stated simple conditions ensuring the rate optimality and efficiency of approximate martingale estimating function-based estimators of the drift and diffusion parameters α and β. As mentioned in Section 1.2.1, he also argued that the theory of approximate martingale estimating functions covers a considerable number of other estimators proposed in the literature on continuous diffusions. In light of these considerations, it is our belief that an in-depth study of the asymptotic theory of approximate martingale estimating functions for jump-diffusions is not only justified, but imperative, and could contribute valuable information to the field of parametric estimation for diffusions with jumps. The overall goal of Chapter 3 is to provide preliminary findings in this regard.

More specifically, our primary objective in Chapter 3 is as follows: We aim to establish existence, uniqueness, and asymptotic distribution results for consistent, approximate martingale estimating function-based estimators of θ_0 in the general model (1.3.1), under the present observation scheme.

Subsequently, we focus on the sub-model (1.3.2), for which Shimizu and Yoshida (2006) obtained efficient estimators. Our secondary objective in Chapter 3 is the following: We strive to give conditions on the approximate martingale estimating functions, which ensure rate optimality and efficiency of estimators of the drift-jump and diffusion parameters. Unlike the efficient contrast function of Shimizu and Yoshida (2006), approximate martingale estimating functions are not a priori designed to discriminate between observed increments with jumps and those without. We expect this distinguishing mechanism to be an inherent feature of the rate optimality and efficiency conditions, to the extent that it is necessary.

1.3.2 Overview of Main Results

Let $G_n(\theta)$ be an approximate martingale estimating function as described in Section 1.2.1, given by the deterministic $\mathbb{R}^d$-valued function $g(t, y, x; \theta)$.

The first main contribution of Chapter 3 is regarding the general model (1.3.1), within the framework described in Section 1.3.1. Theorem 3.3.2 establishes existence, uniqueness, and asymptotic distribution results for consistent G_n-estimators of the true, unknown parameter θ_0. In general terms, omitting details and regularity conditions, the theorem may be summarised as follows.

Theorem. *Suppose that the appropriate assumptions hold. Then,*

(i) there exists a consistent G_n-estimator $\hat{\theta}_n$. In any compact, convex set $K \subseteq \Theta$ containing θ_0 in its interior, the estimator is unique with probability going to one.

(ii) for any consistent G_n-estimator $\hat{\theta}_n$, it holds that

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, V(\theta_0)) \,. \tag{1.3.4}$$

$V(\theta_0)$ is estimated consistently by $\widehat{V}_n$, yielding the more practically applicable result

$$\sqrt{n\Delta_n}\, \widehat{V}_n^{-1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, I_d) \,,$$

where I_d denotes the $d \times d$ identity matrix.

◇

In particular, (1.3.4) is comparable to the asymptotic results derived by Masuda (2011, 2013) for certain Gaussian quasi-likelihood functions, which fit into the theory of approximate martingale estimating functions. A concrete example is given in Example 3.3.3 of Chapter 3.

We pursue the question of rate optimality and efficiency in three types of sub-models of (1.3.2). The first is assumed to have only an unknown, d-dimensional drift-jump parameter α, the second only an unknown, one-dimensional diffusion parameter β, and the third a two-dimensional drift-jump parameter α and a one-dimensional diffusion parameter β, both unknown. In this connection, it should be noted that G_n-estimators of the drift-jump parameter are already seen to be rate optimal by the general convergence result in (1.3.4), whereas there is room for improvement in the rate of convergence of the diffusion parameters.

Our second main contribution in Chapter 3 consists of the following:[6] For the two classes of sub-models containing diffusion parameters, we give conditions which ensure rate optimality of G_n-estimators of these parameters (Conditions 3.4.10 and 3.4.14), and establish

[6]Under the Conjecture 3.4.4, which is definitely true for the models for which Becheri et al. (2014) and Kohatsu-Higa et al. (2014, 2015) established the local asymptotic normality property.

their limit distributions (Theorems 3.4.11 and 3.4.15). For all three classes of sub-models, we state additional conditions on the approximate martingale estimating functions, under which the estimators are also efficient (Conditions 3.4.6, 3.4.12 and 3.4.16).

The conditions we find extend the conditions given by Sørensen (2010) for rate optimality and efficiency of the drift and diffusion parameters in continuous diffusion models of the form (1.3.2) with $c(x,z;\alpha)\equiv 0$. In the limit $\Delta_n\to 0$, the full sample path of $\mathbf{X}$ is (hypothetically) observed. Then, in general terms, $g(t,y,x;\theta)$ and its derivatives should be thought of as evaluated at $(t,y,x)=(0,X_t,X_{t-})$ instead of $(t,y,x)=(\Delta_n,X_{t_i^n},X_{t_{i-1}^n})$. For continuous diffusions, $X_t=X_{t-}$ at all times. Thus it makes sense intuitively, that rate optimality and efficiency entail conditions on the functions when evaluated at $(0,y,x)=(0,x,x)$, as seen in Sørensen's paper. For jump-diffusions however, $X_t\neq X_{t-}$ at jump times, so it seems reasonable that additional conditions could be needed on the off-diagonal $y\neq x$, as seen in our case.

Let d_1 and d_2 respectively denote the dimension of the drift-jump parameter α and the diffusion parameter β, with $d=d_1+d_2\geq 1$. Define $g^\star=(g_\alpha^\star,g_\beta^\star)$, where $g_\alpha(t,y,x;\theta)$ is $\mathbb{R}^{d_1}$-valued and $g_\beta(t,y,x;\theta)$ is $\mathbb{R}^{d_2}$-valued. In Sørensen (2010)[7], the simple condition $\partial_y g_\beta(0,x,x;\theta)=0$ for all x and θ, ensured rate optimality of estimators of the diffusion parameter. For jump-diffusions, our investigation as described above reveals the following: In order to obtain rate optimality for the diffusion parameter, $g_\beta(0,y,x;\theta)$ and several of its partial derivatives need to vanish at an increased number of points depending on the jump dynamics of the process. Even more when a drift-jump parameter is also present in the model. For certain jump-diffusions, it might be difficult or even impossible to construct g_β so that the rate optimality conditions and e.g., the implied non-degeneracy condition of Theorem 3.4.15, are satisfied simultaneously. The latter entails that $\partial_y^2 g_\beta(0,x,x;\theta)$ does not vanish π_θ-almost surely for any θ in the parameter set, which could easily conflict with the rate optimality condition that, e.g. $g(0,y,x;\theta)$ should vanish for "many" $y\neq x$.

Regarding the supplementary conditions for efficiency, the condition found for the diffusion parameter is identical to that of Sørensen (2010) (and Chapter 2), requiring that

$$\partial_y^2 g_\beta(0,x,x;\theta)=K_\theta^{(2)}\frac{\partial_\beta b^2(x;\beta)}{b^4(x;\beta)}$$

for a non-zero constant $K_\theta^{(2)}$, for all x and θ. The conditions found for the drift-jump parameter are more involved. For example, with $\varphi(x,w;\alpha)$ and $\mathcal{W}(x)$ as described in Section 1.3.1, an efficient choice of g_α should satisfy that for all x and θ,

$$g_\alpha(0,x+w,x;\theta)=K_\theta^{(1)}\frac{\partial_\alpha\varphi(x,w;\alpha)^\star}{\varphi(x,w;\alpha)}\quad\text{and}\quad\partial_y g_\alpha(0,x,x;\theta)=K_\theta^{(1)}\frac{\partial_\alpha a(x;\alpha)^\star}{b^2(x;\beta)},$$

for Lebesgue-almost all $w\in\mathcal{W}(x)$, where $K_\theta^{(1)}$ is a non-zero, possibly θ-dependent constant. In other words, $g_\alpha(0,y,x;\theta)$ should be able to discriminate between pairs $(y,x)=(X_t,X_{t-})$ with $X_t\neq X_{t-}$ and $X_t=X_{t-}$. Whenever $y\neq x$ with $y-x$ the possible size of a

[7] And in Chapter 2 of this thesis.

jump increment, $g_\alpha(0, y, x; \theta)$ is determined by the score function of the jump distribution (and must vanish if α is not actually present in the jump mechanism of $\mathbf{X}$). For $y = x$, $g_\alpha(0, y, x; \theta)$ should behave like an efficient estimating function of the drift parameter in a continuous diffusion, as the second equation is in accordance with the efficiency condition given by Sørensen for said parameter.

1.3.3 Conclusions and Perspectives for Further Research

In Chapter 3, we considered approximate martingale estimating function-based estimation for ergodic diffusions with finite-activity jumps. The processes were assumed to be observed at high frequency over an increasing time interval, with terminal sampling time going to infinity. Existence, uniqueness properties and asymptotic distribution results were established for consistent estimators, in a model with a general, finite-dimensional parameter. Rate optimality and efficiency criteria were motivated by existing results in the literature. Subsequently, conditions were given for rate optimality and efficiency of the estimators in three classes of sub-models, with an unknown drift-jump parameter and/or an unknown diffusion parameter. These conditions were found to extend the pre-existing conditions applicable to continuous diffusions, but imposed considerably stronger requirements on the estimating functions.

It was stated that the overall aim of our study in Chapter 3 was to provide preliminary findings on the topic of asymptotic theory for general approximate martingale estimating functions for jump-diffusions. In our opinion, we succeeded in this respect. First, we proved a general existence and convergence result for consistent estimators, thus confirming that the topic is viable. Secondly, the additional conditions which were provided constitute a starting point for further research. Obvious next steps would be, for example, to determine to what extent it is possible to find rate optimal and efficient approximate martingale estimating functions in the presence of jumps. Furthermore, to construct concrete examples of such functions. In Chapter 3, we briefly discussed how, in certain models, the contrast function (1.3.3) proposed by Shimizu and Yoshida (2006) could perhaps be modified to fit our framework, possibly by weakening our regularity assumptions as well.

On a slightly different note, inspired by the investigation of Masuda (2013) into efficiency loss in connection with the Gaussian quasi-likelihood estimators, it could be beneficial to investigate the efficiency loss associated with more general approximate martingale estimating functions. It might be possible to make use of the knowledge thus obtained, to create efficient or nearly efficient estimating functions.

Moreover, inspired by our simulated example in Chapter 2, it would be useful to study estimators based on general (not necessarily efficient or rate optimal) approximate martingale estimating functions by simulation. This could be, for example, in order to ascertain how well they perform for finite samples, and to determine their practical usefulness.

CHAPTER 2

Efficient Estimation for Diffusions Sampled at High Frequency Over a Fixed Time Interval

Nina Munkholt Jakobsen & Michael Sørensen

Department of Mathematical Sciences
University of Copenhagen

Abstract

This paper considers parametric estimation for univariate diffusion processes, which are observed at high frequency over a fixed time interval. The processes are assumed to solve stochastic differential equations with an unknown parameter present only in the diffusion coefficient. Using approximate martingale estimating functions, we obtain consistent estimators of the parameter, which are rate optimal and under an additional condition, also efficient in the local asymptotic mixed normality sense. When suitably normalised, the estimators converge in distribution to normal variance-mixtures. These limit distributions may be characterised as the product of two independent random variables, one of which is standard normally distributed. The other generally depends on the full path of the diffusion process over the observation time interval, as well as on the true, unknown parameter. Utilising the concept of stable convergence in distribution, we also obtain the more practicable result that when normalised slightly differently, the estimators converge in distribution to a standard normal distribution. An example of an efficient estimating function is given, and it is argued that more may be found in the literature. To exemplify the theory, we perform a small simulation study using two estimating functions, one of them efficient, where we make various graphical comparisons of the asymptotic distributions of the estimators.

2.1 Introduction

Diffusions given by stochastic differential equations find application in a number of fields where they are used to describe phenomena which evolve in continuous time. Some examples include agronomy (Pedersen, 2000), biology (Favetto and Samson, 2010), finance (Cox et al., 1985; De Jong et al., 2001; Merton, 1971; Vasicek, 1977) and neuroscience (Bibbona et al., 2010; Ditlevsen and Lansky, 2006; Picchini et al., 2008).

While the models have continuous-time dynamics, the data is mainly only observable in discrete time, thus creating a demand for statistical methods to analyse such data. With the exception of some simple cases, the likelihood function is not explicitly known, making maximum likelihood estimation somewhat infeasible.

A large variety of alternate estimation procedures have been proposed in the literature. Parametric methods include the following: Maximum likelihood-type estimation using, primarily, Gaussian types of approximations to the likelihood function was considered by Florens-Zmirou (1989), Genon-Catalot (1990), Genon-Catalot and Jacod (1993), Gloter and Sørensen (2009), Jacod (2006), Kessler (1997), Prakasa Rao (1983), Sørensen and Uchida (2003), and Yoshida (1992). Analytical expansions of the transition densities were investigated by Aït-Sahalia (2002, 2008) and Li (2013), while approximations to the score function were studied by Bibby and Sørensen (1995), Jacobsen (2001, 2002), Sørensen (2010), and Uchida (2004). Also, simulation-based likelihood methods were developed by Beskos et al. (2006, 2009), Durham and Gallant (2002), Pedersen (1995), and Roberts and Stramer (2001).

Non-parametric methods have been studied as well, see, e.g. Bandi and Phillips (2003), Comte et al. (2007), Florens-Zmirou (1993), Genon-Catalot et al. (1992), Jacod (2000), and Schmisser (2013). Recently, Papaspiliopoulos et al. (2012), van der Meulen and van Zanten (2013), and van der Meulen et al. (2014) focused on the development of Bayesian non-parametric methods.

This paper concerns parametric estimation in a setup where the diffusion process $\mathbf{X} = (X_t)_{t\geq 0}$ solves a stochastic differential equation of the form

$$dX_t = a(X_t)\,dt + b(X_t;\theta)\,dW_t\,, \tag{2.1.1}$$

where $(W_t)_{t\geq 0}$ is a standard Wiener process. The drift and diffusion coefficients a and b are known, deterministic functions of y and $(y;\theta)$, respectively, and θ is the unknown parameter to be estimated. For ease of exposition, X_t and θ are both assumed to be one-dimensional. At least, the extension of our results to a multivariate parameter is expected to be quite straightforward. For each sample size $n \in \mathbb{N}$, we assume observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$ of $\mathbf{X}$ over the interval $[0,1]$, at discrete, equidistant time-points $t_i^n = i/n$ with $i = 0, 1, \ldots, n$.[1] Asymptotics are considered as $n \to \infty$. The diffusion is said to be sampled at *high frequency*, because the time step $\Delta_n = 1/n$ satisfies that $\Delta_n \to 0$ as $n \to \infty$.

[1]With a slight abuse of terminology, as there are, in fact, $n + 1$ observations.

The choice of the time-interval $[0, 1]$ is not particularly restrictive, results generalise to other compact intervals by suitable rescaling of the drift and diffusion coefficients. No parameter is assumed in the drift coefficient, as such parameters cannot be estimated consistently in the asymptotic scenario under consideration. Here, and in the following, e.g. $\partial_u f$ denotes the (partial) derivative of a function f with respect to the variable u.

It was shown by Dohnal (1987); Gobet (2001) that the local asymptotic mixed normality property holds within this setup, with rate $\sqrt{n}$ and random asymptotic Fisher information

$$\mathcal{I}(\theta_0) = 2\int_0^1 \left(\frac{\partial_\theta b(X_s;\theta_0)}{b(X_s;\theta_0)}\right)^2 ds = \frac{1}{2}\int_0^1 \left(\frac{\partial_\theta b^2(X_s;\theta_0)}{b^2(X_s;\theta_0)}\right)^2 ds\,.$$

In this context, a consistent estimator $\hat{\theta}_n$ of the unknown, true parameter θ_0 is said to be rate optimal if $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to a non-degenerate random variable as $n \to \infty$. Furthermore, the estimator is said to be efficient if the limit may be written on the form $\mathcal{I}(\theta_0)^{-1/2}Z$, where Z follows a standard normal distribution and is independent of $\mathcal{I}(\theta_0)$. These concepts are elaborated in Section 2.2.6.

Estimation in the situation described above was considered by Genon-Catalot and Jacod (1993, 1994), within the framework of a more general model and observation scheme.[2] Genon-Catalot and Jacod (1993) proposed estimators based on a class of constrast functions, which were only allowed to depend on the observations through $b^2(X_{t_{i-1}^n};\theta)$ and $\Delta_n^{-1/2}(X_{t_i^n} - X_{t_{i-1}^n})$. These estimators were shown to be rate optimal, and an example was given of an efficient estimator.

In this paper, we investigate estimators based on the extensive class of approximate martingale estimating functions

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,.$$

The real-valued function $g(t, y, x;\theta)$ satisfies a conditional expectation condition of the form

$$\mathbb{E}_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}) = \Delta_n^\kappa R_\theta(\Delta_n, X_{t_{i-1}^n})$$

for some $\kappa \geq 2$, where the remainder term $R_\theta(t, x)$ on the right-hand side can be controlled as necessary. Estimators are essentially obtained as solutions to the estimating equation $G_n(\theta) = 0$. More precise definitions of the estimating functions and estimators are given in Section 2.2.3.

Estimating functions of the (approximate) martingale type were used by, e.g. Bibby and Sørensen (1995), Jacobsen (2001, 2002), Sørensen (2010) and Uchida (2004), in connection with other models and asymptotic schemes (see also Sørensen (2012)). In particular, the model given by (2.1.1) is a sub-model of that considered by Sørensen (2010), who studied approximate martingale estimating functions for high frequency observations over an

[2]In the following, we disregard the extended setup of Genon-Catalot and Jacod and focus on the interpretation of their results within the model and observation scheme of the present paper.

increasing time interval with terminal sampling time $t_n^n = n\Delta_n \to \infty$ as $n \to \infty$. Not only were simple conditions for rate optimality and efficiency given (there, in a local asymptotic normality sense), it was also argued that the theory of approximate martingale estimating functions covers a large number of other well-performing estimators in the literature.

First, we establish existence and uniqueness results regarding consistent estimators $\hat{\theta}_n$ of the true parameter θ_0, which are based on approximate martingale estimating functions. We show that these estimators are rate optimal, in that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to a normal variance-mixture distribution. The limit distribution may be represented by the product $W(\theta_0)Z$ of independent random variables, where Z follows a standard normal distribution. $W(\theta_0)$ is generally random, and depends on the path of the diffusion process over the time-interval $[0, 1]$.

Normal variance-mixture distributions were also obtained as the asymptotic distributions of the estimators of Genon-Catalot and Jacod (1993). These distributions appear as limit distributions in comparable non-parametric settings as well, e.g. when estimating integrated volatility $\int_0^1 b^2(X_s)\,ds$ (Jacod and Protter, 1998; Mykland and Zhang, 2006) or the squared diffusion coefficient $b^2(x)$ (Florens-Zmirou, 1993; Jacod, 2000).

Rate optimality is ensured by the condition that

$$\partial_y g(0, x, x; \theta) = 0 \tag{2.1.2}$$

for all x in the state space of $\mathbf{X}$, and all parameters θ, where $\partial_y g(0, x, x; \theta)$ denotes the first derivative of $g(0, y, x; \theta)$ with respect to y, evaluated in $y = x$. This was the same condition found for rate optimality of the estimator of the diffusion parameter in Sørensen (2010). It was referred to by Sørensen as *Jacobsen's condition*, as it is one of the conditions for small Δ-optimality in the sense of Jacobsen (2001), for a model with only a diffusion parameter (Jacobsen, 2002). Jacobsen (2002) considered near-efficiency of martingale estimating function-based estimators using discrete observations, with a fixed distance Δ close to 0 between observation times. The asymptotic covariance matrix of the estimators was expanded in powers of Δ in the limit $\Delta \to 0$, and, loosely put, small Δ-optimal estimators were those which minimised the leading term of this expansion.

For some models, $W(\theta_0)$ does not depend on $(X_s)_{s\in[0,1]}$. In these cases, $W(\theta_0)$ is deterministic, making $W(\theta_0)Z$ a zero-mean normal distribution with variance $W(\theta_0)^2$. Otherwise, however, due to its dependence on $(X_s)_{s\in[0,1]}$, the limit distribution is not particularly useful for statistical applications, such as constructing confidence intervals and test statistics. Therefore, we construct $\widehat{W}_n$, a function of $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$, which converges in probability to $W(\theta_0)$. Taking into account that there is actually stable convergence in distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ towards $W(\theta_0)Z$, we are then able to derive the more practically applicable result that $\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0)$ converges in distribution to a standard normal distribution.

The additional condition that

$$\partial_y^2 g(0, x, x; \theta) = K_\theta \frac{\partial_\theta b^2(x; \theta)}{b^4(x; \theta)} \tag{2.1.3}$$

for all x in the state space of $\mathbf{X}$, and all parameters θ, ensures efficiency of our estimators in the local asymptotic mixed normality-framework. (Here, $K_\theta \neq 0$ is a possibly θ-dependent constant.) This condition was also obtained by Sørensen (2010) for efficiency of the diffusion parameter-estimator, and is identical to another condition given by Jacobsen (2002) for small Δ-optimality. The approximate martingale estimating function given by

$$g(t, y, x; \theta) = \frac{\partial_\theta b^2(x;\theta)}{b^4(x;\theta)} \Big((y-x)^2 - t b^2(x;\theta) \Big) \tag{2.1.4}$$

satisfies (2.1.2) and (2.1.3), and corresponds to the contrast function shown to be efficient by Genon-Catalot and Jacod (1993, Theorem 5). By the overlap between conditions, examples of approximate martingale estimating functions satisfying our rate optimality and efficiency conditions may also be found in the papers of Jacobsen (2002) and Sørensen (2010).

To exemplify the theory, we perform a small simulation study based on a model which satisfies our conditions, and for which the limit distribution of the estimators is an actual normal variance-mixture (and not merely a normal distribution). Using two estimating functions, one of them given by (2.1.4) and therefore efficient, we make various graphical comparisons of the asymptotic distributions of the estimators. In accordance with the theoretical considerations, the efficient estimator is seen to have preferable properties.

The assumptions made in this paper are similar to those of Sørensen (2010), although here, ergodicity of $\mathbf{X}$ is not needed to obtain results of the law of large numbers-type. To some extent, we use similar methods of proof as well, e.g. convergence in probability is shown after the expansion of relevant conditional moments in powers of Δ_n. However, due to the differences in the respective asymptotic schemes, higher-order expansions than in the work of Sørensen are sometimes needed here. Furthermore, in order to establish (stable) convergence in distribution in the current paper, a more complicated central limit theorem is required than that used by Sørensen. Finally, while convergence in distribution was sufficient for the asymptotic scenario considered by Sørensen, in our case, due to the randomness of $W(\theta_0)$, we need to entertain the stronger concept of *stable* convergence in distribution, in order to obtain practically applicable convergence results.

The rest of this paper is structured as follows: Section 2.2 presents definitions, notation and terminology used throughout the paper, as well as the main assumptions on the diffusion process and the approximate martingale estimating functions. Section 2.3 states and discusses our main results, including the simulation example. Section 2.4 contains main lemmas used to prove the main theorem, the proof of the main theorem, and the proofs of the main lemmas. Appendix 2.A consists of auxiliary results, some of them with proofs, while Appendix 2.B summarises some important theorems from the literature, without proofs.

2.2 Preliminaries

Section 2.2.1 serves to introduce some notation associated with the diffusion process and the observation scheme under consideration. In Section 2.2.2, a notation and terminology regarding the concept of polynomial growth is established for subsequent use. Section 2.2.3 contains formal definitions of approximate martingale estimating functions and their corresponding estimators. Section 2.2.4 introduces the main assumptions on the diffusion process (Assumption 2.2.4) and the estimating function (Assumption 2.2.5). In Section 2.2.5, notation pertaining to the (infinitesimal) generator of the diffusion process is established, and some useful technical results expressed in terms of the generator are discussed. Finally, in Section 2.2.6, the concept of local asymptotic mixed normality is defined very briefly, and the accompanying notions of rate optimality and efficiency, as adopted in this paper, are elaborated on.

2.2.1 Model and Observations

Let $(\Omega, \mathcal{F})$ be a measurable space which supports a real-valued random variable U, and an independent standard Wiener process $\mathbf{W} = (W_t)_{t\geq 0}$. Let $(\mathcal{F}_t)_{t\geq 0}$ denote the filtration generated by U and $\mathbf{W}$, and let $(\mathbb{P}_\theta)_{\theta\in\Theta}$ be a family of probability measures on $(\Omega, \mathcal{F})$. The one-dimensional parameter set Θ is assumed to contain the true parameter θ_0.

Consider the stochastic differential equation

$$dX_t = a(X_t)\, dt + b(X_t;\theta)\, dW_t\,, \quad X_0 = U\,, \tag{2.2.1}$$

for $\theta \in \Theta$. X_t is assumed to take its values in an open, not necessarily bounded interval $\mathcal{X} \subseteq \mathbb{R}$, and the drift and diffusion coefficients, $a : \mathcal{X} \to \mathbb{R}$ and $b : \mathcal{X} \times \Theta \to \mathbb{R}$ respectively, are assumed to be known, deterministic functions.

Let $t_i^n = i\Delta_n$ with $\Delta_n = 1/n$ for $i \in \mathbb{N}_0$, $n \in \mathbb{N}$. For each $n \in \mathbb{N}$, $\mathbf{X}$ is assumed to be sampled at times t_i^n, $i = 0, 1, \ldots, n$, yielding the observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$. Let $\mathcal{G}_{n,i}$ denote the σ-algebra generated by $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_i^n})$, with $\mathcal{G}_n = \mathcal{G}_{n,n}$. For purely theoretical reasons, observations of $\mathbf{X}$ at times t_i^n for $i > n$ occasionally come into play as well.

2.2.2 Polynomial Growth

In the following, to avoid cumbersome notation, C denotes a generic, strictly positive, real-valued constant. Often, the notation C_u is used to emphasise that the constant depends on u in some unspecified manner, where u may be e.g. a number, a set of parameters or both. It is important to note that, for example, in an expression of the form $C_u(1 + |x|^{C_u})$, the factor C_u and the exponent C_u need not be equal. Generic constants C_u often depend (implicitly) on the unknown parameter θ_0, but never on the sample size n.

Definition 2.2.1. A function $f : [0,1] \times \mathcal{X}^2 \times \Theta \to \mathbb{R}$ is of polynomial growth in x and y, uniformly for $t \in [0,1]$ and θ in compact, convex sets, if for each compact, convex set $K \subseteq \Theta$ there exist constants $C_K > 0$ such that

$$\sup_{t\in[0,1],\,\theta\in K} |f(t,y,x;\theta)| \leq C_K(1 + |x|^{C_K} + |y|^{C_K})$$

for $x, y \in \mathcal{X}$.

$C^{\text{pol}}_{p,q,r}([0,1]\times\mathcal{X}^2\times\Theta)$ denotes the class of continuous, real-valued functions $f(t,y,x;\theta)$ which satisfy that

(i) f and the mixed partial derivatives $\partial_t^i\partial_y^j\partial_\theta^k f(t,y,x;\theta)$, $i=0,\ldots,p$, $j=0,\ldots,q$ and $k=0,\ldots,r$ exist and are continuous on $[0,1]\times\mathcal{X}^2\times\Theta$.

(ii) f and the mixed partial derivatives from (i) are of polynomial growth in x and y, uniformly for $t\in[0,1]$ and θ in compact, convex sets.

Similarly, the classes $C^{\text{pol}}_{p,r}([0,1]\times\mathcal{X}\times\Theta)$, $C^{\text{pol}}_{q,r}(\mathcal{X}^2\times\Theta)$, $C^{\text{pol}}_{q,r}(\mathcal{X}\times\Theta)$ and $C^{\text{pol}}_{q}(\mathcal{X})$ are defined for functions of the form $f(t,x;\theta)$, $f(y,x;\theta)$, $f(y;\theta)$ and $f(y)$, respectively. ⋄

Note that in Definition 2.2.1, differentiability of f with respect to x is never required, and that for functions not depending on t (respectively θ), the "uniformly for t" ("uniformly for θ") part of the definition becomes superfluous.

For the duration of this paper, $R(t,y,x;\theta)$ denotes a generic, real-valued function defined on $[0,1]\times\mathcal{X}^2\times\Theta$, which is of polynomial growth in x and y uniformly for $t\in[0,1]$ and θ in compact, convex sets. $R(t,y,x;\theta)$ may depend (implicitly) on θ_0. $R(t,x;\theta)$, $R(y,x;\theta)$ and $R(t,x)$ are defined correspondingly. The notation $R_\lambda(t,x;\theta)$ indicates that $R(t,x;\theta)$ also depends on $\lambda\in\Theta$ in an unspecified way. In particular, $R_\theta(t,x,\theta)=R_\theta(t,x)$.

2.2.3 Approximate Martingale Estimating Functions

Let $\mathbb{E}_\theta$ denote expectation under $\mathbb{P}_\theta$. In this paper, (approximate) martingale estimating functions, along the lines of those specified by e.g. Sørensen (2012), are defined as follows:

Definition 2.2.2. Let $g(t,y,x;\theta)$ be a real-valued function defined on $[0,1]\times\mathcal{X}^2\times\Theta$. Suppose the existence of a constant $\kappa\geq 2$, such that for all $n\in\mathbb{N}$, $i=1,\ldots,n$, $\theta\in\Theta$,

$$\mathbb{E}_\theta\left(g(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)=\Delta_n^\kappa R_\theta(\Delta_n,X_{t_{i-1}^n})\,. \tag{2.2.2}$$

Then, the function

$$G_n(\theta)=\sum_{i=1}^n g(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta) \tag{2.2.3}$$

is called an *approximate martingale estimating function*. In particular, when (2.2.2) is satisfied with $R_\theta(t,x)\equiv 0$, (2.2.3) is referred to as a *martingale estimating function*. ⋄

By the Markov property of $\mathbf{X}$, it is seen that when $R_\theta(t,x)\equiv 0$, $(G_{n,i})_{1\leq i\leq n}$ defined by

$$G_{n,i}(\theta)=\sum_{j=1}^i g(\Delta_n,X_{t_j^n},X_{t_{j-1}^n};\theta)$$

is a zero-mean, real-valued $(\mathcal{G}_{n,i})_{1\leq i\leq n}$-martingale under $\mathbb{P}_\theta$ for each $n\in\mathbb{N}$, thus giving rise to the terminology in Definition 2.2.2. However, when not ambiguous, approximate martingale estimating functions may sometimes just be referred to as *estimating functions* in the following. An approximate martingale estimating function is essentially an approximation to the score function of the observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$, conditional on $X_{t_0^n}$, which itself is a martingale.

A G_n*-estimator* $\hat{\theta}_n$, that is, an estimator based on the approximate martingale estimating function $G_n(\theta)$, is essentially obtained as a solution to the estimating equation $G_n(\theta) = 0$. A more precise definition, based on the definitions of Jacod and Sørensen (2012, Definition 2.1) and Sørensen (2012, Defintion 1.57), is given in Definition 2.2.3.

Formally, an approximate martingale estimating function may be considered a function of both $\theta \in \Theta$ and $\omega \in \Omega$, while a G_n-estimator may be considered a function of ω. For the purpose of the following definition, it is convenient to make this dependence explicit and write $G_n(\theta, \omega)$ and $\hat{\theta}_n(\omega)$.

Definition 2.2.3. Let $G_n(\theta, \omega)$ be an approximate martingale estimating function as defined in Definition 2.2.2. Put $\Theta_\infty = \Theta \cup \{\infty\}$ and let

$$D_n = \{\omega \in \Omega \mid G_n(\theta, \omega) = 0 \text{ has at least one solution } \theta \in \Theta\}.$$

A G_n*-estimator* $\hat{\theta}_n(\omega)$ is any $\mathcal{G}_n$-measurable function $\Omega \to \Theta_\infty$ which satisfies that for $\mathbb{P}_{\theta_0}$-almost all ω, $\hat{\theta}_n(\omega) \in \Theta$ and $G_n(\hat{\theta}_n(\omega), \omega) = 0$ if $\omega \in D_n$, and $\hat{\theta}_n(\omega) = \infty$ if $\omega \notin D_n$. $\diamond$

For any $M_n \neq 0$, which may depend on e.g. Δ_n, $G_n(\theta)$ and $M_n G_n(\theta)$ yield identical estimators of θ. The estimating functions $G_n(\theta)$ and $M_n G_n(\theta)$ are referred to as *versions* of each other. For any given estimating function, it is sufficient that there exists a version of the function which satisfies the assumptions of this paper, in order to draw conclusions about the resulting estimators.

2.2.4 Assumptions

Assumption 2.2.4. *The parameter set Θ is a non-empty, open, not necessarily bounded subset of $\mathbb{R}$, which contains the true parameter θ_0. The continuous, $(\mathcal{F}_t)_{t\geq 0}$-adapted Markov process $\mathbf{X} = (X_t)_{t\geq 0}$ solves a stochastic differential equation of the form (2.2.1), the coefficients of which satisfy that*

$$a(y) \in C_6^{pol}(\mathcal{X}) \quad \text{and} \quad b(y;\theta) \in C_{6,2}^{pol}(\mathcal{X} \times \Theta) .$$

The following holds for all $\theta \in \Theta$.

(i) For all $y \in \mathcal{X}$, $b^2(y;\theta) > 0$.

(ii) There exists a real-valued constant $C_\theta > 0$ such that for all $x, y \in \mathcal{X}$,

$$|a(x) - a(y)| + |b(x;\theta) - b(y;\theta)| \leq C_\theta |x - y| .$$

(iii) There exists a real-valued constant $C_\theta > 0$ such that for all $y \in \mathcal{X}$,

$$|a(y)| + |b(y;\theta)| \leq C_\theta (1 + |y|) .$$

(iv) For all $m \in \mathbb{N}$,

$$\sup_{t\in[0,\infty)} \mathbb{E}_\theta\left(|X_t|^m\right) < \infty\,.$$

◇

Assumptions 2.2.4.(ii) and (iii), known as the *global Lipschitz* and *linear growth* conditions, ensure that **X** is well-defined. By these assumptions, for each $\theta \in \Theta$, there exists a unique, $(\mathcal{F}_t)_{t\geq 0}$-adapted, non-exploding solution to (2.2.1) with continuous sample paths $t \mapsto X_t(\omega)$, which is a Markov process. For use in the following, observe that under $\mathbb{P}_\theta$, X_t may be written as

$$X_t = X_0 + \int_0^t a(X_s)\, ds + \int_0^t b(X_s;\theta)\, dW_s\,.$$

Assumption 2.2.4 is very similar to the corresponding Condition 2.1 of Sørensen (2010). However, an important difference is that in the current paper, **X** is not required to be ergodic. Here, law of large numbers-type results are obtained by what is, in essence, the convergence of Riemann sums.

Assumption 2.2.5. *The function $g(t,y,x;\theta)$ satisfies that*

$$g(t,y,x;\theta) \in C^{pol}_{3,8,2}([0,1]\times\mathcal{X}^2\times\Theta)\,,$$

and defines an approximate martingale estimating function $G_n(\theta)$ as prescribed by Definition 2.2.2. In particular,

(i) for some constant $\kappa \geq 2$,

$$\mathbb{E}_\theta\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^\kappa R_\theta(\Delta_n, X_{t_{i-1}^n})$$

for all $n \in \mathbb{N}$, $i = 1,\ldots,n$ and $\theta \in \Theta$.

Furthermore, the following holds for all $\theta \in \Theta$.

(ii) For all $x \in \mathcal{X}$, $\partial_y g(0,x,x;\theta) = 0$.

(iii) The expansion

$$\begin{aligned} g(\Delta,y,x;\theta) &= g(0,y,x;\theta) + \Delta g^{(1)}(y,x;\theta) + \tfrac{1}{2}\Delta^2 g^{(2)}(y,x;\theta) + \tfrac{1}{6}\Delta^3 g^{(3)}(y,x;\theta) \\ &\quad + \Delta^4 R(\Delta,y,x;\theta) \end{aligned} \tag{2.2.4}$$

holds for all $\Delta \in [0,1]$ and $x, y \in \mathcal{X}$, where $g^{(j)}(y,x;\theta)$ denotes the jth partial derivative of $g(t,y,x;\theta)$ with respect to t, evaluated in $t = 0$.

◇

Assumption 2.2.5.(ii) is referred to by Sørensen (2010) as *Jacobsen's condition*, as it is one of the conditions of small Δ-optimality in the sense of Jacobsen (2001), for a model with only a diffusion parameter (Jacobsen, 2002). The assumption ensures rate optimality of the estimators in this paper, and, similarly, of the estimators of the diffusion parameter in Sørensen's article.

The assumptions of polynomial growth serve to simplify the exposition and proofs, and could be relaxed. For example, we make use of the fact that due to Assumption 2.2.4.(iv), measurable transformations $f(y, x)$ of $(X_{t_i^n}, X_{t_{i-1}^n})$, which are of polynomial growth in x and y, have finite moments. Instead, at the expense of readability, we could simply have assumed the existence of the moments necessary for our results.

2.2.5 The Infinitesimal Generator

For parameters $\lambda \in \Theta$ and functions $f(y) \in C_2^{\text{pol}}(\mathcal{X})$, define the *(infinitesimal) generator* $\mathcal{L}_\lambda$, through its action on $f(y)$, as

$$\mathcal{L}_\lambda f(y) = a(y)\partial_y f(y) + \tfrac{1}{2}b^2(y;\lambda)\partial_y^2 f(y)\,.$$

More generally, for $f(t, y, x, \theta) \in C_{0,2,0,0}^{\text{pol}}([0,1] \times \mathcal{X}^2 \times \Theta)$, let

$$\mathcal{L}_\lambda f(t, y, x; \theta) = a(y)\partial_y f(t, y, x; \theta) + \tfrac{1}{2}b^2(y;\lambda)\partial_y^2 f(t, y, x; \theta)\,. \tag{2.2.5}$$

Often, the notation $\mathcal{L}_\lambda f(t, y, x; \theta) = \mathcal{L}_\lambda(f(t;\theta))(y, x)$ is used, so e.g. $\mathcal{L}_\lambda(f(0;\theta))(x, x)$ means $\mathcal{L}_\lambda f(0, y, x; \theta)$ evaluated in $y = x$. Whenever well-defined, $\mathcal{L}_\lambda^2 f$ is to be understood as $\mathcal{L}_\lambda(\mathcal{L}_\lambda f)$, and similarly $\mathcal{L}_\lambda^k f = \mathcal{L}_\lambda(\mathcal{L}_\lambda^{k-1} f)$ for $k \in \mathbb{N}$, with $\mathcal{L}_\lambda^0 f = f$.

The infinitesimal generator notation is particularly useful for expressing the result of the following Lemma 2.2.6.

Lemma 2.2.6. *Suppose that Assumption 2.2.4 holds, and that for some $k \in \mathbb{N}_0$,*

$$a(y) \in C_{2k}^{pol}(\mathcal{X})\,, \quad b(y;\theta) \in C_{2k,0}^{pol}(\mathcal{X} \times \Theta) \quad \text{and} \quad f(y, x; \theta) \in C_{2(k+1),0}^{pol}(\mathcal{X}^2 \times \Theta)\,.$$

Then, for $0 \leq t \leq t + \Delta \leq 1$ and $\lambda \in \Theta$,

$$\begin{aligned}&\mathbb{E}_\lambda\left(f(X_{t+\Delta}, X_t; \theta) \mid X_t\right)\\ &= \sum_{i=0}^{k} \frac{\Delta^i}{i!}\mathcal{L}_\lambda^i f(X_t, X_t; \theta) + \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E}_\lambda\left(\mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t\right) du_{k+1} \cdots du_1\end{aligned}$$

where, furthermore,

$$\int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E}_\lambda\left(\mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t\right) du_{k+1} \cdots du_1 = \Delta^{k+1} R_\lambda(\Delta, X_t; \theta)\,.$$

◇

The expansion of the conditional expectation in powers of Δ in the first part of the lemma corresponds to the expression of Florens-Zmirou (1989, Lemma 1) (or Dacunha-Castelle and Florens-Zmirou (1986, Lemma 4), after the correction of a small typo). It may be proven by induction on k using Itô's formula, see, for example, the proof of Sørensen (2012, Lemma 1.10). As seen in the proof of Kessler (1997, Lemma 1), the characterisation of the remainder term follows by applying Corollary 2.A.5 to $\mathcal{L}_\lambda^{k+1} f$.[3]

Assumption 2.2.4 ensures that $a(y) \in C_{2k}^{\text{pol}}(\mathcal{X})$ and $b(y;\theta) \in C_{2k,0}^{\text{pol}}(\mathcal{X}\times\Theta)$ for $k = 0, 1, 2, 3$, so when Lemma 2.2.6 is used for these values of k (as is done in this paper), the assumptions on a and b are automatically satisfied.

In addition to its application in proofs presented in this paper, Lemma 2.2.6 is, together with Assumption 2.2.5.(i), key to proving Lemma 2.2.7, which reveals two important properties of the approximate martingale estimating functions. Lemma 2.2.7 corresponds to Lemma 2.3 of Sørensen (2010), to which we refer for further details on the proof.

Lemma 2.2.7. *Suppose that Assumptions 2.2.4 and 2.2.5 hold. Then*

$$g(0, x, x; \theta) = 0 \quad \textit{and} \quad g^{(1)}(x, x; \theta) = -\mathcal{L}_\theta(g(0, \theta))(x, x)$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$. ◇

In concrete examples, Lemma 2.2.6 is also useful for verifying Assumption 2.2.5.(i), a fundamental property of approximate martingale estimating functions, and, conversely, it can be used to create such estimating functions as well.

2.2.6 Local Asymptotic Mixed Normality

Note that in this paper, all convergence in probability and convergence in distribution, denoted $\xrightarrow{\mathcal{P}}$ and $\xrightarrow{\mathcal{D}}$ respectively, is assumed to be under $\mathbb{P}_{\theta_0}$ as $n \to \infty$.

Local asymptotic mixed normality was introduced by Jeganathan (1982), and is discussed in e.g. Jacod (2010) and Le Cam and Yang (2000, Chapter 6). Below, the local asymptotic mixed normality property is defined in a univariate setting, along the lines of the definition presented by Jacod (2010, Section 3.2).

Recall that $\mathcal{G}_n$ is the σ-algebra generated by the observations $(X_{t_0^n}, X_{t_1^n}, \dots, X_{t_n^n})$, and let $\mathbb{P}_\theta^n$ denote the restriction of $\mathbb{P}_\theta$ to $\mathcal{G}_n$. Define the likelihood ratios $Q_n(\lambda;\theta) = \log(d\mathbb{P}_\lambda^n / d\mathbb{P}_\theta^n)$.

Definition 2.2.8. Suppose that there exist sequences $R_n(\theta_0)$ and $\mathcal{I}_n(\theta_0)$ of $\mathcal{G}_n$-measurable random variables with $\mathbb{P}_{\theta_0}(\mathcal{I}_n(\theta_0) > 0) = 1$, and a deterministic sequence δ_n of strictly positive real numbers with $\delta_n \to \infty$ as $n \to \infty$, such that for all $u \in \mathbb{R}$,

$$Q_n\left(\theta_0 + \frac{u}{\delta_n}, \theta_0\right) - uR_n(\theta_0) + \frac{u^2}{2}\mathcal{I}_n(\theta_0) \xrightarrow{\mathcal{P}} 0$$

[3] The last part of Section 3.A.4 in Chapter 3 of this thesis is dedicated to the proof of Lemma 3.2.8, a jump-diffusion counterpart to Lemma 2.2.6. Itself being a modification and extension of the proofs presented by Flachs (2011); Sørensen (2012) in the case of ergodic continuous diffusions, this proof is easily converted to a proof of Lemma 2.2.6.

and

$$(R_n(\theta_0), \mathcal{I}_n(\theta_0)) \xrightarrow{\mathcal{D}} (\mathcal{I}(\theta_0)^{1/2}Z, \mathcal{I}(\theta_0)),$$

where Z is standard normally distributed, $\mathbb{P}_{\theta_0}(\mathcal{I}(\theta_0) > 0) = 1$, and Z and $\mathcal{I}(\theta_0)$ are independent. Then, the statistical model $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ is *locally asymptotically mixed normal* at θ_0, with rate δ_n and random Fisher information $\mathcal{I}(\theta_0)$. ◇

$\mathcal{I}(\theta_0)$ is generally random, and may be interpreted as a measure of how well θ_0 can be estimated, based on, in the current setting, a particular realisation of $(X_t)_{t\in[0,1]}$.

Rate Optimality and Efficiency

In the context of local asymptotic mixed normality, the definitions of rate optimality and efficiency are quite straightforward.

Definition 2.2.9. Suppose that the model $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ for $\mathbf{X}$ is locally asymptotically mixed normal at θ_0, with rate δ_n and random Fisher information $\mathcal{I}(\theta_0)$. Then, a sequence of estimators $\hat{\theta}_n$ is *rate optimal* if $\delta_n(\hat{\theta}_n - \theta_0)$ converges in distribution to a non-degenerate limit under $\mathbb{P}_{\theta_0}$ as $n \to \infty$. Additionally, the sequence is *efficient* if the limit is the normal variance-mixture $\mathcal{I}(\theta_0)^{-1/2}Z$, with Z standard normally distributed and independent of $\mathcal{I}(\theta_0)$. ◇

Loosely put, δ_n is the "fastest possible" rate of convergence in distribution, the best rate at which θ_0 can be estimated. See e.g. Jacod (2010) for further details. Also, it was shown by Jeganathan (1982) that if the local asymptotic mixed normality property is satisfied, and $\hat{\theta}_n$ is a rate optimal but not necessarily efficient estimator of θ_0 with $\delta_n(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} L(\theta_0)$, the following holds under certain further assumptions. Conditionally on $\mathcal{I}(\theta_0)$, the distribution of $L(\theta_0)$ is a convolution of the zero-mean normal distribution with variance $\mathcal{I}(\theta_0)^{-1}$ and some other distribution. Loosely put, the distribution of $L(\theta_0)$ is more spread out than the specified normal distribution. Let $\mathbb{V}_\theta$ denote variance under $\mathbb{P}_\theta$. If the relevant quantities exist, $\mathbb{V}_{\theta_0}(L(\theta_0) \mid \mathcal{I}(\theta_0)) \geq \mathcal{I}(\theta_0)^{-1}$, implying that also unconditionally,

$$\mathbb{V}_{\theta_0}(L(\theta_0)) = \mathbb{E}_{\theta_0}\left(\mathbb{V}_{\theta_0}(L(\theta_0) \mid \mathcal{I}(\theta_0))\right) + \mathbb{V}_{\theta_0}\left(\mathbb{E}_{\theta_0}(L(\theta_0) \mid \mathcal{I}(\theta_0))\right) \geq \mathbb{E}_{\theta_0}(\mathcal{I}(\theta_0)^{-1}).$$

If the estimator is efficient, the lower bounds are achieved, i.e. the estimator has the smallest possible variance, both when conditioned on $\mathcal{I}(\theta_0)$, and unconditionally (the latter only if the unconditional variance exists).

2.3 Main Results

In Section 2.3.1, the main theorem of this paper, Theorem 2.3.2, is presented. The theorem establishes existence, uniqueness and asymptotic distribution results for rate optimal estimators of θ_0 based on approximate martingale estimating functions. In Section 2.3.2, a condition is stated, which ensures that the rate optimal estimators found in Theorem 2.3.2 are also efficient, and efficient estimators are discussed in some further detail. Section 2.3.3 contains an example of the theory, in the form of a small simulation study.

2.3.1 Main Theorem

Assumption 2.3.1 is the final assumption needed for the main theorem, Theorem 2.3.2. The notation A, B and C corresponds to the notation of Lemma 2.4.1, and is used in the proof of Theorem 2.3.2 as well.

Assumption 2.3.1. *The following holds $\mathbb{P}_\theta$-almost surely for all $\theta \in \Theta$.*

(i) For all $\lambda \neq \theta$,

$$A(\lambda, \theta) = \tfrac{1}{2} \int_0^1 (b^2(X_s; \theta) - b^2(X_s; \lambda)) \partial_y^2 g(0, X_s, X_s; \lambda)\, ds \neq 0\,.$$

(ii) Furthermore,

$$B(\theta; \theta) = -\tfrac{1}{2} \int_0^1 \partial_\theta b^2(X_s; \theta) \partial_y^2 g(0, X_s, X_s; \theta)\, ds \neq 0\,,$$

(iii) and

$$C(\theta; \theta) = \tfrac{1}{2} \int_0^1 b^4(X_s; \theta) \partial_y^2 g(0, X_s, X_s; \theta)^2\, ds \neq 0\,.$$

◇

Assumption 2.3.1 can be difficult to check in practice, seeing that it involves the full sample path of $\mathbf{X}$ over the interval $[0, 1]$. It requires, in particular, that for all $\theta \in \Theta$, with $\mathbb{P}_\theta$-probability one, $t \mapsto b^2(X_t; \theta) - b^2(X_t; \lambda)$ is not Lebesgue-almost surely zero when $\lambda \neq \theta$ (Genon-Catalot and Jacod, 1993, Hypothesis H4). As also noted by Genon-Catalot and Jacod, this requirement holds true (by the continuity of the function) if, for example, $X_0 = U$ is distributed according to ε_{x_0}, the degenerate probability measure with point mass in x_0, and $b^2(x_0; \theta) \neq b^2(x_0; \lambda)$ for all $\theta \neq \lambda$.

In Section 2.3.2, it becomes clear that for an efficient estimating function, Assumption 2.3.1 reduces to conditions on $\mathbf{X}$, more specifically, conditions involving only the squared diffusion coefficient $b^2(x; \theta)$ and its derivative $\partial_\theta b^2(x; \theta)$, with no further conditions on the estimating function.

Theorem 2.3.2 is the main theorem presented in this paper. Recalling that Dohnal (1987) and Gobet (2001) showed that the local asymptotic mixed normality property holds at θ_0 with rate $\sqrt{n}$ and random Fisher information

$$\mathcal{I}(\theta_0) = 2 \int_0^1 \left(\frac{\partial_\theta b(X_s; \theta_0)}{b(X_s; \theta_0)} \right)^2 ds = \frac{1}{2} \int_0^1 \left(\frac{\partial_\theta b^2(X_s; \theta_0)}{b^2(X_s; \theta_0)} \right)^2 ds \tag{2.3.1}$$

within the current framework, Theorem 2.3.2 establishes rate optimal G_n-estimators of θ_0, based on approximate martingale estimating functions. (See Definition 2.2.3 for the definition of a G_n-estimator.)

Theorem 2.3.2. *Suppose that Assumptions 2.2.4, 2.2.5 and 2.3.1 hold. Then,*

(i) there exists a consistent G_n-estimator $\hat{\theta}_n$. Choose any compact, convex set $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$, where $\operatorname{int} K$ denotes the interior of K. Then, the consistent G_n-estimator $\hat{\theta}_n$ is eventually unique in K, in the sense that for any G_n-estimator $\tilde{\theta}_n$ with $\mathbb{P}_{\theta_0}(\tilde{\theta}_n \in K) \to 1$ as $n \to \infty$, it holds that $\mathbb{P}_{\theta_0}(\hat{\theta}_n \neq \tilde{\theta}_n) \to 0$ as $n \to \infty$.

(ii) for any consistent G_n-estimator $\hat{\theta}_n$, it holds that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} W(\theta_0)Z\,. \tag{2.3.2}$$

The limit distribution is a normal variance-mixture, where Z is standard normally distributed, and independent of $W(\theta_0)$ given by

$$W(\theta_0) = \frac{\left(\int_0^1 \tfrac{1}{2} b^4(X_s;\theta_0)\partial_y^2 g(0,X_s,X_s;\theta_0)^2\,ds\right)^{1/2}}{\int_0^1 \tfrac{1}{2}\partial_\theta b^2(X_s;\theta_0)\partial_y^2 g(0,X_s,X_s;\theta_0)\,ds}\,. \tag{2.3.3}$$

(iii) for any consistent G_n-estimator $\hat{\theta}_n$,

$$\widehat{W}_n = -\frac{\left(\frac{1}{\Delta_n}\sum_{i=1}^n g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\theta}_n)\right)^{1/2}}{\sum_{i=1}^n \partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\theta}_n)} \tag{2.3.4}$$

satisfies that $\widehat{W}_n \xrightarrow{\mathcal{P}} W(\theta_0)$, and

$$\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} Z\,,$$

where Z is standard normally distributed.

◇

Observe that the limit distribution in Theorem 2.3.2.(ii) generally depends on not only the unknown parameter θ_0, but also on the concrete realisation of the sample path $t \mapsto X_t$ over $[0,1]$, which is only partially observed. In contrast, Theorem 2.3.2.(iii) yields a limit distribution which is of more use in practical applications. The proof of Theorem 2.3.2 is given in Section 2.4.2.

2.3.2 Efficiency

Under the assumptions of Theorem 2.3.2, the additional condition for efficiency of a consistent G_n-estimator is given in Assumption 2.3.3. It is identical to the condition for efficiency of estimators of the diffusion parameter given by Sørensen (2010, Condition 1.2) and, like Assumption 2.2.5.(ii), one of the conditions for small Δ-optimality found by Jacobsen (2002).

Assumption 2.3.3. *Suppose that for each $\theta \in \Theta$, there exists a constant $K_\theta \neq 0$ such that for all $x \in \mathcal{X}$,*

$$\partial_y^2 g(0, x, x; \theta) = K_\theta \frac{\partial_\theta b^2(x;\theta)}{b^4(x;\theta)} .$$

◇

Within the framework considered here, Definition 2.2.9 prescribes efficiency of a G_n-estimator $\hat{\theta}_n$ when (2.3.2) holds with $W(\theta_0) = \mathcal{I}(\theta_0)^{-1/2}$, and $\mathcal{I}(\theta_0)$ is given by (2.3.1). Thus, Corollary 2.3.4 may easily be verified.

Corollary 2.3.4. *Suppose that the assumptions of Theorem 2.3.2 and Assumption 2.3.3 hold. Then, any consistent G_n-estimator is also efficient.* ◇

It was noted in Section 2.2.3 that not necessarily all versions of a particular estimating function satisfy the conditions of this paper, even though they may be used to obtain the same estimator. Thus, an estimating function is said to be efficient, if there exists a version which satisfies the conditions of Corollary 2.3.4. The same goes for rate optimality.

Under suitable regularity conditions on the diffusion coefficient b, the function

$$\bar{g}(t, y, x; \theta) = \frac{\partial_\theta b^2(x;\theta)}{b^4(x;\theta)} \left((y-x)^2 - t b^2(x;\theta) \right) \tag{2.3.5}$$

yields one example of an efficient estimating function $G_n(\theta) = \sum_{i=1}^n \bar{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$. The approximate martingale property, Assumption 2.2.5.(i), can be verified by the help of Lemma 2.2.6.

When adapted to the current framework, the contrast functions investigated by Genon-Catalot and Jacod (1993) have the form

$$U_n(\theta) = \frac{1}{n} \sum_{i=1}^n f\left(b^2(X_{t_{i-1}^n};\theta), \Delta_n^{-1/2}(X_{t_i^n} - X_{t_{i-1}^n}) \right) ,$$

for functions $f(v, w)$ satisfying certain conditions. For the contrast function identified as efficient by Theorem 5 of Genon-Catalot and Jacod, $f(v, w) = \log v + w^2/v$. Using that $\Delta_n = 1/n$, it is then seen that their efficient contrast function is of the form $\bar{U}_n(\theta) = \sum_{i=1}^n \bar{u}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$ with

$$\bar{u}(t, y, x; \theta) = t \log b^2(x;\theta) + (y-x)^2 / b^2(x;\theta)$$

and $\partial_\theta \bar{u}(t, y, x; \theta) = -\bar{g}(t, y, x; \theta)$. In other words, it corresponds to a version of the efficient approximate martingale estimating function given by (2.3.5).

A problem of considerable practical interest is how to construct estimating functions that are (rate optimal and) efficient, i.e. estimating functions satisfying Assumptions 2.2.5.(ii) and 2.3.3. Being the same as the conditions for small Δ-optimality in a model with only a diffusion parameter (Jacobsen, 2002), the assumptions are, for example, satisfied by martingale estimating functions constructed by Jacobsen.

As discussed by Sørensen (2010), the rate optimality and efficiency conditions are also satisfied by Godambe-Heyde optimal approximate martingale estimating functions. Consider martingale estimating functions of the form

$$g(t, y, x; \theta) = a(x, t; \theta)\left(f(y; \theta) - \phi_\theta^t f(x; \theta)\right) ,$$

where $\phi_\theta^t f(x;\theta) = \mathbb{E}_\theta(f(X_t;\theta) \mid X_0 = x)$, and suppose that f satisfies appropriate conditions. Let $\bar{a}$ be the weight function for which the estimating function is optimal in the sense of Godambe and Heyde, see e.g. Godambe and Heyde (1987); Heyde (1997) or Sørensen (2012, Section 1.11). It follows by an argument analogous to the proof of Theorem 4.5 in Sørensen (2010) that the estimating function with

$$g(t, y, x; \theta) = t\bar{a}(x, t; \theta)[f(y; \theta) - \phi_\theta^t f(x; \theta)]$$

satisfies Assumptions 2.2.5.(ii) and 2.3.3, and is thus rate optimal and efficient. As there is a simple formula for $\bar{a}$ (see Section 1.11.1 of Sørensen (2012)), this provides a way of constructing a large number of efficient estimating functions. The result also holds if $\phi_\theta^t f(x;\theta)$ and the conditional moments in the formula for $\bar{a}$ are approximated suitably by the help of Lemma 2.2.6.

Remark 2.3.5. Suppose for a moment that the diffusion coefficient of (2.2.1) may be parametrised such that $b^2(x;\theta) = h(x)k(\theta)$ for suitable, strictly positive functions h and k, with Assumption 2.2.4 satisfied. This holds true for e.g. a number of Pearson diffusions, including the (stationary) Ornstein-Uhlenbeck and square root processes. (See Forman and Sørensen (2008) for more on Pearson diffusions.) Then $\partial_\theta b^2(x;\theta) = h(x)\partial_\theta k(\theta)$, yielding $\mathcal{I}(\theta_0) = \partial_\theta k(\theta_0)^2/(2k^2(\theta_0))$. In this case, under the assumptions of Corollary 2.3.4, an efficient G_n-estimator $\hat{\theta}_n$ satisfies that

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} Y$$

where Y is normally distributed with mean zero and variance $2k^2(\theta_0)/\partial_\theta k(\theta_0)^2$. That is, for certain "nice" models (2.2.1), the limit distribution of the efficient estimators is simply a zero-mean normal distribution with variance depending on θ_0, rather than a normal variance-mixture depending on θ_0 *and* $(X_t)_{t\in[0,1]}$. ∘

As mentioned in Section 2.3.1, when Assumption 2.3.3 is satisfied, Assumption 2.3.1 reduces to an assumption involving only $b^2(x;\theta)$ and $\partial_\theta b^2(x;\theta)$. In particular, in the special case described in Remark 2.3.5, for an efficient estimating function, Assumption 2.3.1 is satisfied when e.g. $\partial_\theta k(\theta) > 0$ or $\partial_\theta k(\theta) < 0$.

2.3.3 Example: Simulations

This section contains an example of the theory discussed in the previous sections. An efficient and an inefficient estimating function are compared by simulation, and the model under investigation is chosen so that the limit distributions of the consistent estimators obtained by Theorem 2.3.2.(ii) are non-degenerate normal variance-mixtures, in the sense that they do not trivialise to normal distributions.

Model and Estimating Functions

Consider the stochastic differential equation

$$dX_t = -\alpha X_t\,dt + (\theta + \beta X_t^2)^{-1/2}\,dW_t \tag{2.3.6}$$

where $\alpha, \beta > 0$ are known constants and $\theta \in (0, \infty)$ is an unknown parameter. Then $\mathbf{X}$ is ergodic, and the invariant probability measure has density proportional to

$$\mu_\theta(x) = \exp\left(-\alpha\theta x^2 - \tfrac{1}{2}\alpha\beta x^4\right)\left(\theta + \beta x^2\right), \quad x \in \mathbb{R}, \tag{2.3.7}$$

with respect to Lebesgue measure. It may be verified that when $\mathbf{X}$ is stationary, the process satisfies Assumption 2.2.4. Two estimating functions are considered, $G_n(\theta)$ and $H_n(\theta)$ given by

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \quad \text{and} \quad H_n(\theta) = \sum_{i=1}^n h(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$$

where

$$\begin{aligned} g(t, y, x; \theta) &= (y - x)^2 - (\theta + \beta x^2)^{-1} t \\ h(t, y, x; \theta) &= (\theta + \beta x^2)^{10}(y - x)^2 - (\theta + \beta x^2)^9 t\,. \end{aligned}$$

Both g and h satisfy Assumptions 2.2.5 and 2.3.1, and g may be recognised as the efficient function (2.3.5), while h is not efficient.

Let $W_G(\theta_0)$ and $W_H(\theta_0)$ be given by (2.3.3) for the respective estimating functions, that is

$$W_G(\theta_0) = \left(\frac{1}{2}\int_0^1 \frac{1}{(\theta_0 + \beta X_s^2)^2}\,ds\right)^{-1/2} \quad \text{and} \quad W_H(\theta_0) = \frac{\left(\int_0^1 2(\theta_0 + \beta X_s^2)^{18}\,ds\right)^{1/2}}{\int_0^1 (\theta_0 + \beta X_s^2)^8\,ds}\,. \tag{2.3.8}$$

Simulations

In this section, numerical calculations and simulations were done in R 3.1.2 (R Core Team, 2014). First, $m = 10^4$ trajectories of the process $\mathbf{X}$ given by (2.3.6) were simulated over the time-interval $[0, 1]$ with $\alpha = 2, \beta = 1$ and $\theta_0 = 1$, each with sample size $n = 10^4$. These simulations were performed using the R-package *sde* (Iacus, 2014). For each trajectory, the initial value X_0 was obtained from the invariant distribution of $\mathbf{X}$ by inverse transform sampling, using a quantile function based on (2.3.7), and calculated by numerical procedures in R. For $n = 10^3$ and $n = 10^4$, let $\hat\theta_{G,n}$ and $\hat\theta_{H,n}$ denote estimates of θ_0 obtained by solving the equations $G_n(\theta) = 0$ and $H_n(\theta) = 0$ numerically, on the interval $[0.01, 1, 99]$. Using these estimates, $\widehat{W}_{G,n}$ and $\widehat{W}_{H,n}$ are calculated by (2.3.4).[4]

[4]For $n = 10^3$, $\hat\theta_{H,n}$ and thus also $\widehat{W}_{H,n}$, could not be computed for 44 of the $m = 10^4$ sample paths. For $n = 10^4$, and for the efficient estimator $\hat\theta_{G,n}$ there were no problems.

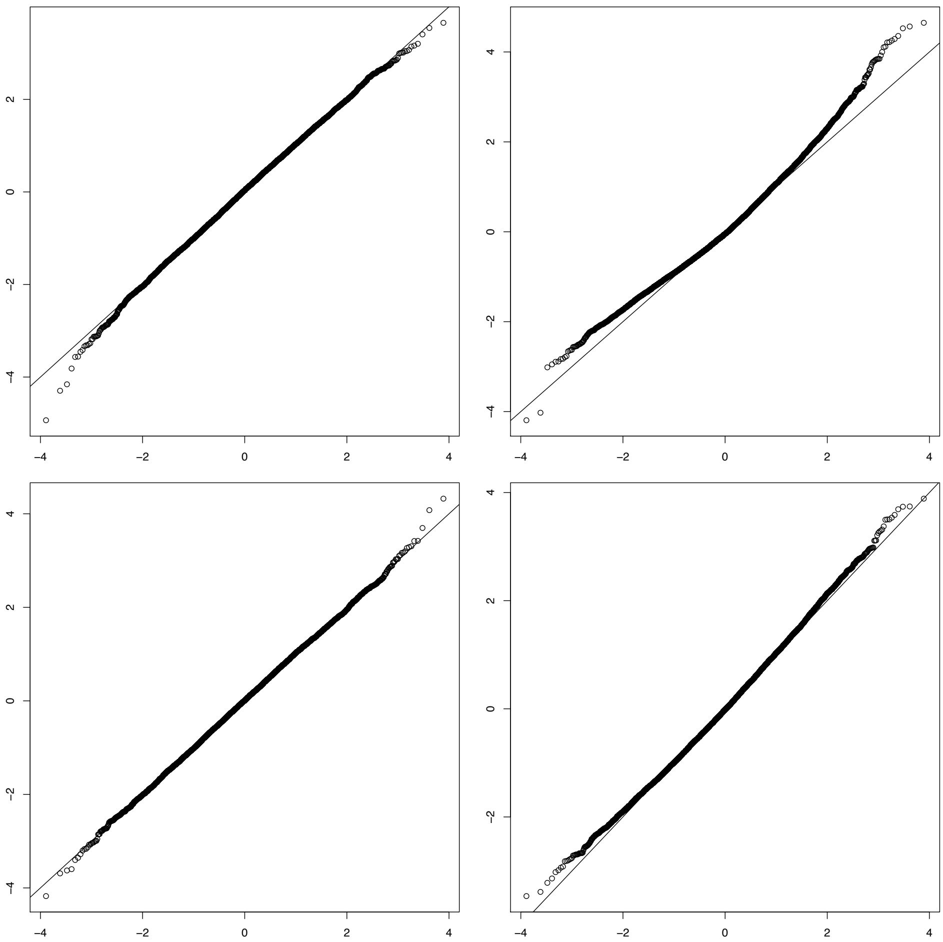

Figure 2.1: QQ-plots comparing $\widehat{Z}_{G,n}$ (left) and $\widehat{Z}_{H,n}$ (right) to the $\mathcal{N}(0,1)$ distribution for $n = 10^3$ (above) and $n = 10^4$ (below).

Figure 2.1 shows QQ-plots of

$$\widehat{Z}_{G,n} = \sqrt{n}\,\widehat{W}_{G,n}^{-1}(\hat{\theta}_{G,n} - \theta_0) \quad \text{and} \quad \widehat{Z}_{H,n} = \sqrt{n}\,\widehat{W}_{H,n}^{-1}(\hat{\theta}_{H,n} - \theta_0)\,,$$

compared with a standard normal distribution, for $n = 10^3$ and $n = 10^4$ respectively. These QQ-plots suggest that at least in the current example, as n goes to infinity, the asymptotic distribution in Theorem 2.3.2.(iii) becomes a good approximation faster in the efficient case than in the inefficient case.

Inserting $\theta_0 = 1$ into (2.3.8), the intergrals in these expressions may be approximated by Riemann sums, using each of the simulated trajectories of $\mathbf{X}$ (with $n = 10^4$ for maximal accuracy). This method yields a second set of approximations $\widetilde{W}_G$ and $\widetilde{W}_H$ to the realisations of the random variables $W_G(\theta_0)$ and $W_H(\theta_0)$, presumed to be more accurate than $\widehat{W}_{G,10^4}$ and $\widehat{W}_{H,10^4}$ as they utilise the true parameter. The *density* function in R was used (with default arguments) to compute an approximation to the densities of $W_G(\theta_0)$ and $W_H(\theta_0)$, using the approximate realisations $\widetilde{W}_G$ and $\widetilde{W}_H$.

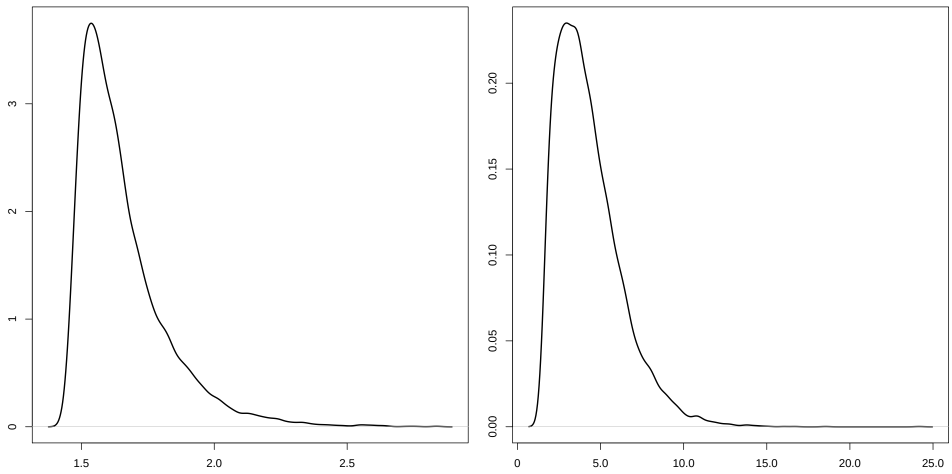

Figure 2.2: Approximation to the densities of $W_G(\theta_0)$ (left) and $W_H(\theta_0)$ (right) based on $\widetilde{W}_G$ and $\widetilde{W}_H$.

It is seen from Figure 2.2 that the distribution of $W_H(\theta_0)$ is much more spread out than the distribution of $W_G(\theta_0)$. This corresponds well to the limit distribution in Theorem 2.3.2.(ii) being more spread out in the inefficient case than in the efficient case. Along the same lines, Figure 2.3 shows similarly computed densities based on $\sqrt{n}(\hat{\theta}_{G,n}-\theta_0)$ and $\sqrt{n}(\hat{\theta}_{H,n}-\theta_0)$ for $n = 10^4$, which may be considered approximations to the densities of the normal variance-mixture limit distributions in Theorem 2.3.2.(ii). These plots also illustrate that the limit distribution of the inefficient estimator is more spread out than that of the efficient estimator.

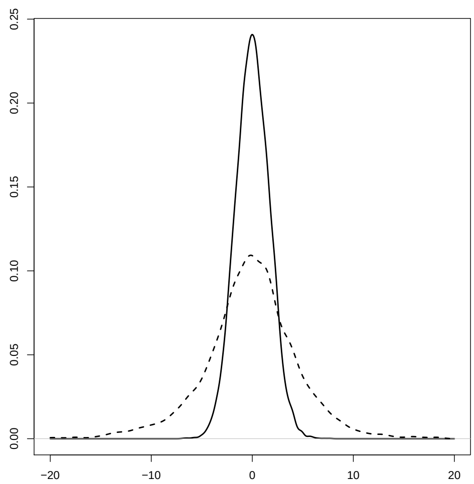

Figure 2.3: Estimated densities of $\sqrt{n}(\hat{\theta}_{G,n} - \theta_0)$ (solid curve) and $\sqrt{n}(\hat{\theta}_{H,n} - \theta_0)$ (dashed curve) for $n = 10^4$.

2.4 Proofs

Section 2.4.1 states several lemmas needed to prove Theorem 2.3.2, and a brief definition of stable convergence in distribution is given. Theorem 2.3.2 is proved in Section 2.4.2. Section 2.4.3 contains the proofs of the three main lemmas from Section 2.4.1.

2.4.1 Main Lemmas

In order to prove Theorem 2.3.2, the lemmas presented in this section are utilised, together with results on the existence, uniqueness and convergence of G_n-estimators from Jacod and Sørensen (2012), and Sørensen (2012, Section 1.10). Proofs of the main Lemmas 2.4.1, 2.4.2 and 2.4.4 are given in Section 2.4.3. In particular, a stable limit theorem, Theorem IX.7.28 of Jacod and Shiryaev (2003), is used to prove the stable convergence in distribution in Lemma 2.4.4.

For convenience, the applicable theorems of Jacod, Shiryaev and Sørensen are briefly summarised in Appendix 2.B, in a simplified form, tailored specifically to fit the framework and needs of the current paper. Stable convergence in distribution is defined, also very briefly and with minimum technicality, prior to the presentation of Lemma 2.4.4.

Lemma 2.4.1. *Suppose that Assumptions 2.2.4 and 2.2.5 hold. For $\theta \in \Theta$, let*

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$$

$$G_n^{sq}(\theta) = \frac{1}{\Delta_n} \sum_{i=1}^{n} g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$$

and

$$
\begin{aligned}
A(\theta;\theta_0) &= \tfrac{1}{2}\int_0^1 \left(b^2(X_s;\theta_0) - b^2(X_s;\theta)\right)\partial_y^2 g(0,X_s,X_s;\theta)\,ds \\
B(\theta;\theta_0) &= \tfrac{1}{2}\int_0^1 \left(b^2(X_s;\theta_0) - b^2(X_s;\theta)\right)\partial_y^2\partial_\theta g(0,X_s,X_s;\theta)\,ds \\
&\quad - \tfrac{1}{2}\int_0^1 \partial_\theta b^2(X_s;\theta)\partial_y^2 g(0,X_s,X_s;\theta)\,ds \\
C(\theta;\theta_0) &= \tfrac{1}{2}\int_0^1 \left(b^4(X_s;\theta_0) + \tfrac{1}{2}\left(b^2(X_s;\theta_0) - b^2(X_s;\theta)\right)^2\right)\partial_y^2 g(0,X_s,X_s;\theta)^2\,ds\,.
\end{aligned}
$$

Then,

(i) the mappings $\theta \mapsto A(\theta;\theta_0)$*,* $\theta \mapsto B(\theta;\theta_0)$ *and* $\theta \mapsto C(\theta;\theta_0)$ *are continuous on* Θ *(*$\mathbb{P}_{\theta_0}$*-almost surely) with* $A(\theta_0;\theta_0) = 0$ *and* $\partial_\theta A(\theta;\theta_0) = B(\theta;\theta_0)$*.*

(ii) for all $t > 0$*,*

$$
\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{[nt]}\left|\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right)\right| \xrightarrow{\mathcal{P}} 0 \tag{2.4.1}
$$

$$
\frac{1}{\Delta_n}\sum_{i=1}^{[nt]}\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right)^2 \xrightarrow{\mathcal{P}} 0 \tag{2.4.2}
$$

$$
\frac{1}{\Delta_n^2}\sum_{i=1}^{[nt]}\mathbb{E}_{\theta_0}\left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right) \xrightarrow{\mathcal{P}} 0 \tag{2.4.3}
$$

and

$$
\frac{1}{\Delta_n}\sum_{i=1}^{[nt]}\mathbb{E}_{\theta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right) \xrightarrow{\mathcal{P}} \tfrac{1}{2}\int_0^t b^4(X_s;\theta_0)\partial_y^2 g(0,X_s,X_s;\theta_0)^2\,ds\,. \tag{2.4.4}
$$

(iii) for all compact, convex sets $K \subseteq \Theta$*,*

$$
\sup_{\theta\in K}\left|G_n(\theta) - A(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0
$$

$$
\sup_{\theta\in K}\left|\partial_\theta G_n(\theta) - B(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0
$$

$$
\sup_{\theta\in K}\left|G_n^{sq}(\theta) - C(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0\,.
$$

(iv) for any consistent estimator $\hat\theta_n$ *of* θ_0*,*

$$
\partial_\theta G_n(\hat\theta_n) \xrightarrow{\mathcal{P}} B(\theta_0;\theta_0) \quad \text{and} \quad G_n^{sq}(\hat\theta_n) \xrightarrow{\mathcal{P}} C(\theta_0;\theta_0)\,.
$$

◇

Lemma 2.4.2. *Suppose that Assumptions 2.2.4 and 2.2.5 hold. Then, for all $t > 0$,*

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} 0 \,. \tag{2.4.5}$$

◇

Stable Convergence in Distribution

The results in this paper make use of the concept stable convergence in distribution, as introduced by Rényi (1963) and discussed in the works of e.g. Aldous and Eagleson (1978), Hall and Heyde (1980), Jacod (1997), Jacod and Shiryaev (2003), and in the survey article of Podolskij and Vetter (2010). Stable convergence in distribution implies, in particular, convergence in distribution. The implication is evident from the definition below. Here, the *random elements* Y_n are either real-valued random variables or continuous, univariate stochastic processes, but, as seen in the references, the definition easily generalises. Definition 2.4.3 is a slightly modified version of Definition 1 in the paper of Podolskij and Vetter.

Definition 2.4.3. Let $(Y_n)_{n\in\mathbb{N}}$ be a sequence of random elements defined on $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$, and Y a random element defined on an extension $(\Omega', \mathcal{F}', \mathbb{P}'_\theta)$ of $(\Omega, \mathcal{F}, \mathbb{P}_\theta)$. The sequence $(Y_n)_{n\in\mathbb{N}}$ converges *stably in distribution* to Y under $\mathbb{P}_\theta$ as $n \to \infty$ if, and only if, for all bounded, continuous, real-valued functions h, and all bounded, $\mathcal{F}$-measurable, real-valued random variables Z,

$$\mathbb{E}_\theta(h(Y_n)Z) \to \mathbb{E}'_\theta(h(Y)Z)$$

as $n \to \infty$, $\mathbb{E}'_\theta$ denoting expectation under $\mathbb{P}'_\theta$. ◇

Lemma 2.4.4. *Suppose that Assumptions 2.2.4 and 2.2.5 hold. Let*

$$Y_{n,t} = \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \,.$$

The sequence of processes $(\mathbf{Y}_n)_{n\in\mathbb{N}}$ given by $\mathbf{Y}_n = (Y_{n,t})_{t\geq 0}$ converges stably in distribution under $\mathbb{P}_{\theta_0}$ to the process $\mathbf{Y} = (Y_t)_{t\geq 0}$ given by

$$Y_t = \tfrac{1}{\sqrt{2}} \int_0^t b^2(X_s; \theta_0) \partial_y^2 g(0, X_s, X_s; \theta_0) \, dB_s \,.$$

$\mathbf{B} = (B_s)_{s\geq 0}$ denotes a standard Wiener process, which is defined on a filtered extension $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t\geq 0}, P'_{\theta_0})$ of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P_{\theta_0})$, and is independent of U and $\mathbf{W}$. ◇

As of now, stable convergence in distribution under $\mathbb{P}_{\theta_0}$ as $n \to \infty$ is denoted by $\xrightarrow{\mathcal{D}_{st}}$.

Lemma 2.4.5 and 2.4.6 summarise properties of stable convergence in distribution which will be made use of in the proof of Theorem 2.3.2.

Lemma 2.4.5. *Let V_n, W_n, V and W be real-valued random variables, V_n, W_n, W defined on $(\Omega, \mathcal{F}, \mathbb{P}_{\theta_0})$, and V defined on an extension $(\Omega', \mathcal{F}', \mathbb{P}'_{\theta_0})$ of $(\Omega, \mathcal{F}, \mathbb{P}_{\theta_0})$. Suppose that $V_n \xrightarrow{\mathcal{D}_{st}} V$ and $W_n \xrightarrow{\mathcal{P}} W$. Then,*

(i) $(V_n, W_n) \xrightarrow{\mathcal{D}_{st}} (V, W)$.

(ii) for $g : \mathbb{R}^2 \to \mathbb{R}$ continuous on $C \subseteq \mathbb{R}^2$ with $\mathbb{P}'_{\theta_0}((V, W) \in C) = 1$,

$$g(V_n, W_n) \xrightarrow{\mathcal{D}_{st}} g(V, W)\,.$$

◇

Lemma 2.4.6. *Let $\mathbf{Y}_n = (Y_{n,t})_{t\geq 0}$ be a sequence of continuous, adapted, real-valued stochastic processes defined on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_{\theta_0})$, and let $\mathbf{Y} = (Y_t)_{t\geq 0}$ be defined on a filtered extension $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t\geq 0}, \mathbb{P}'_{\theta_0})$ of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_{\theta_0})$. If $\mathbf{Y}_n \xrightarrow{\mathcal{D}_{st}} \mathbf{Y}$ then, for fixed $t_0 \geq 0$,*

$$Y_{n,t_0} \xrightarrow{\mathcal{D}_{st}} Y_{t_0}\,.$$

◇

Lemma 2.4.5.(i), see, e.g. (2.3) in Jacod (1997), may be viewed as an improvement of the result that $V_n \xrightarrow{\mathcal{D}} V$ and $W_n \xrightarrow{\mathcal{P}} w$, $w \in \mathbb{R}$ constant, implies $(V_n, W_n) \xrightarrow{\mathcal{D}} (V, w)$, and is key to obtaining Theorem 2.3.2.(iii). Lemma 2.4.5.(ii) and Lemma 2.4.6, on the other hand, correspond to well-known properties of convergence in distribution, and follow easily from Definition 2.4.3 (and Lemma 2.4.5.(i), when showing (ii)).

2.4.2 Proof of Main Theorem

This section contains the proof of Theorem 2.3.2.

Proof of Theorem 2.3.2. Let any compact, convex subset $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$ be given, and recall that

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)\,.$$

By Lemma 2.4.1.(i) and (iii), and Assumption 2.3.1.(ii),

$$G_n(\theta_0) \xrightarrow{\mathcal{P}} 0 \quad \text{and} \quad \sup_{\theta \in K} |\partial_\theta G_n(\theta) - B(\theta, \theta_0)| \xrightarrow{\mathcal{P}} 0 \tag{2.4.6}$$

with $B(\theta_0; \theta_0) \neq 0$, so $G_n(\theta)$ satisfies the conditions of Theorem 2.B.2 (Sørensen, 2012, Theorem 1.58).

Now, we show (2.B.1) of Theorem 2.B.3 (Sørensen, 2012, Theorem 1.59). Let $\varepsilon > 0$ be given, and let $\bar{B}_\varepsilon(\theta_0)$ and $B_\varepsilon(\theta_0)$, respectively, denote closed and open balls in $\mathbb{R}$ with radius

$\varepsilon > 0$, centered at θ_0. The compact set $K\backslash B_\varepsilon(\theta_0)$ does not contain θ_0, and so, by Assumption 2.3.1.(i), $A(\theta, \theta_0) \neq 0$ for all $\theta \in K\backslash B_\varepsilon(\theta_0)$ with probability one under $\mathbb{P}_{\theta_0}$.

Because

$$\inf_{\theta \in K\backslash \bar{B}_\varepsilon(\theta_0)} |A(\theta, \theta_0)| \geq \inf_{\theta \in K\backslash B_\varepsilon(\theta_0)} |A(\theta, \theta_0)| > 0$$

$\mathbb{P}_{\theta_0}$-almost surely, by the continuity of $\theta \mapsto A(\theta, \theta_0)$, it follows that

$$\mathbb{P}_{\theta_0}\left(\inf_{\theta \in K\backslash \bar{B}_\varepsilon(\theta_0)} |A(\theta, \theta_0)| > 0 \right) = 1 \,.$$

Consequently, by Theorem 2.B.3, for any G_n-estimator $\tilde{\theta}_n$,

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K\backslash \bar{B}_\varepsilon(\theta_0) \right) \to 0 \quad \text{as} \quad n \to \infty \,. \tag{2.4.7}$$

for any $\varepsilon > 0$.

By Theorem 2.B.2, there exists a consistent G_n-estimator $\hat{\theta}_n$, which is eventually unique, in the sense that if $\bar{\theta}_n$ is another consistent G_n-estimator, then

$$\mathbb{P}_{\theta_0}\left(\hat{\theta}_n \neq \bar{\theta}_n \right) \to 0 \quad \text{as} \quad n \to \infty \,. \tag{2.4.8}$$

Suppose that $\tilde{\theta}_n$ is any G_n-estimator which satisfies that

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K \right) \to 1 \quad \text{as} \quad n \to \infty \,. \tag{2.4.9}$$

By (2.4.7) also

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K^c \cup \bar{B}_\varepsilon(\theta_0) \right) \to 1 \quad \text{as} \quad n \to \infty \,, \tag{2.4.10}$$

and combining (2.4.9) and (2.4.10), it follows that $\tilde{\theta}_n$ is consistent. Using (2.4.8), Theorem 2.3.2.(i) follows.

To prove Theorem 2.3.2.(ii), recall that $\Delta_n = 1/n$, and observe that by Lemma 2.4.4 (and Lemma 2.4.6),

$$\sqrt{n} G_n(\theta_0) \xrightarrow{\mathcal{D}_{st}} S(\theta_0) \tag{2.4.11}$$

where

$$S(\theta_0) = \int_0^1 \tfrac{1}{\sqrt{2}} b^2(X_s; \theta_0) \partial_y^2 g(0, X_s, X_s; \theta_0)\, dB_s \,,$$

and $\mathbf{B}$ is a standard Wiener process, independent of U and $\mathbf{W}$. As $\mathbf{X}$ is then also independent of $\mathbf{B}$, $S(\theta_0)$ is equal in distribution to $C(\theta_0; \theta_0)^{1/2} Z$, where Z is standard normally distributed and independent of $(X_t)_{t \geq 0}$. Note that by Assumption 2.3.1.(iii), the distribution of $C(\theta_0; \theta_0)^{1/2} Z$ is non-degenerate.

Let $\hat{\theta}_n$ be a consistent G_n-estimator. By (2.4.6), (2.4.11) and properties of stable convergence (Lemma 2.4.5.(i)),

$$\begin{pmatrix} \sqrt{n}G_n(\theta_0) \\ \partial_\theta G_n(\theta_0) \end{pmatrix} \xrightarrow{\mathcal{D}_{st}} \begin{pmatrix} S(\theta_0) \\ B(\theta_0;\theta_0) \end{pmatrix}.$$

Recalling that stable convergence in distribution implies weak convergence, an application of Theorem 2.B.4 (Sørensen, 2012, Theorem 1.60) yields

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} -B(\theta_0,\theta_0)^{-1}S(\theta_0). \tag{2.4.12}$$

The limit is equal in distribution to $W(\theta_0)Z$, where $W(\theta_0) = -B(\theta_0,\theta_0)^{-1}C(\theta_0;\theta_0)^{1/2}$ and Z is standard normally distributed and independent of $W(\theta_0)$. This completes the proof of Theorem 2.3.2.(ii).

Finally, Lemma 2.B.5 (Jacod and Sørensen, 2012, Lemma 2.14) is used to write

$$\sqrt{n}(\hat{\theta}_n - \theta_0) = -B(\theta_0;\theta_0)^{-1}\sqrt{n}G_n(\theta_0) + \sqrt{n}|\hat{\theta}_n - \theta_0|\varepsilon_n(\theta_0),$$

where the last term goes to zero in probability under $\mathbb{P}_{\theta_0}$. By the stable continuous mapping theorem (Lemma 2.4.5.(ii)), (2.4.12) holds with stable convergence in distribution as well. Lemma 2.4.1.(iv) may be used to conclude that $\widehat{W}_n \xrightarrow{\mathcal{P}} W(\theta_0)$, so Theorem 2.3.2.(iii) follows from the stable version of (2.4.12), by application of Lemma 2.4.5. □

2.4.3 Proofs of Main Lemmas

This section contains the proofs of Lemmas 2.4.1, 2.4.2 and 2.4.4 from Section 2.4.1. A number of technical results are utilised in the proofs, these results are summarised in Section 2.A, some of them with a proof.

Proof of Lemma 2.4.1. First, note that for any $f(x;\theta) \in C^{\text{pol}}_{0,0}(\mathcal{X}\times\Theta)$, $\lambda \in \Theta$ and compact, convex set $K \subseteq \Theta$ with $\lambda \in \operatorname{int} K$, there exist constants $C_K > 0$ such that

$$|f(X_s;\theta)| \leq C_K(1 + |X_s|^{C_K})$$

for all $s \in [0,1]$ and $\theta \in \operatorname{int} K$. With probability one under $\mathbb{P}_{\theta_0}$, for fixed ω, the integral

$$\int_0^1 C_K(1 + |X_s(\omega)|^{C_K})\,ds$$

is simply the integral of a continuous function over $[0,1]$ and therefore finite. Using this method of constructing Lebesgue-integrable upper bounds, Lemma 2.4.1.(i) follows by the usual results for continuity and differentiability of functions given by integrals.

In the rest of this proof, Lemma 2.A.3 and (2.A.7) are repeatedly made use of without reference. First, inserting $\theta = \theta_0$ into (2.A.1), it is seen that

$$\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{[nt]}\left|\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right)\right| = \Delta_n^{3/2}\sum_{i=1}^{[nt]}R(\Delta_n, X_{t_{i-1}^n};\theta_0) \xrightarrow{\mathcal{P}} 0$$

$$\frac{1}{\Delta_n}\sum_{i=1}^{[nt]}\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\mid X_{t_{i-1}^n}\right)^2 = \Delta_n^3\sum_{i=1}^{[nt]}R(\Delta_n, X_{t_{i-1}^n};\theta_0)\xrightarrow{\mathcal{P}}0\,,$$

proving (2.4.1) and (2.4.2). Furthermore, using (2.A.1) and (2.A.3),

$$\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)\xrightarrow{\mathcal{P}}A(\theta;\theta_0)$$
$$\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)\xrightarrow{\mathcal{P}}0\,,$$

so it follows from Lemma 2.A.1 that point-wise for $\theta\in\Theta$,

$$G_n(\theta)-A(\theta;\theta_0)\xrightarrow{\mathcal{P}}0\,. \tag{2.4.13}$$

Using (2.A.3) and (2.A.5),

$$\frac{1}{\Delta_n}\sum_{i=1}^{[nt]}\mathbb{E}_{\theta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)$$
$$\xrightarrow{\mathcal{P}}\tfrac{1}{2}\int_0^t\left(b^4(X_s;\theta_0)+\tfrac{1}{2}\left(b^2(X_s;\theta_0)-b^2(X_s;\theta)\right)^2\right)\partial_y^2 g(0,X_s,X_s;\theta)^2\,ds$$

and

$$\frac{1}{\Delta_n^2}\sum_{i=1}^{[nt]}\mathbb{E}_{\theta_0}\left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)\xrightarrow{\mathcal{P}}0\,,$$

completing the proof of Lemma 2.4.1.(ii) when $\theta=\theta_0$ is inserted, and yielding

$$G_n^{sq}(\theta)-C(\theta;\theta_0)\xrightarrow{\mathcal{P}}0 \tag{2.4.14}$$

point-wise for $\theta\in\Theta$ by Lemma 2.A.1, when $t=1$ is inserted. Also, using (2.A.2) and (2.A.4),

$$\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)\xrightarrow{\mathcal{P}}B(\theta;\theta_0)$$
$$\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left((\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta))^2\mid X_{t_{i-1}^n}\right)\xrightarrow{\mathcal{P}}0\,.$$

Thus, by Lemma 2.A.1, also

$$\partial_\theta G_n(\theta)-B(\theta;\theta_0)\xrightarrow{\mathcal{P}}0\,, \tag{2.4.15}$$

point-wise for $\theta\in\Theta$. Finally, recall that $\partial_y^j g(0,x,x;\theta)=0$ for $j=0,1$. Then, using Lemmas 2.A.7 and 2.A.8, it follows that for each $m\in\mathbb{N}$ and compact, convex subset $K\subseteq\Theta$, there exist constants $C_{m,K}>0$ such that for all $\theta,\theta'\in K$ and $n\in\mathbb{N}$,

$$\begin{aligned}\mathbb{E}_{\theta_0}|(G_n(\theta)-A(\theta;\theta_0))-(G_n(\theta')-A(\theta';\theta_0))|^{2m}&\leq C_{m,K}\,|\theta-\theta'|^{2m}\\ \mathbb{E}_{\theta_0}|(\partial_\theta G_n(\theta)-B(\theta;\theta_0))-(\partial_\theta G_n(\theta')-B(\theta';\theta_0))|^{2m}&\leq C_{m,K}\,|\theta-\theta'|^{2m}\\ \mathbb{E}_{\theta_0}|(G_n^{sq}(\theta)-C(\theta;\theta_0))-(G_n^{sq}(\theta')-C(\theta';\theta_0))|^{2m}&\leq C_{m,K}\,|\theta-\theta'|^{2m}\,.\end{aligned} \tag{2.4.16}$$

By Lemma 2.4.1.(i), the functions $\theta \mapsto G_n(\theta) - A(\theta;\theta_0)$, $\theta \mapsto \partial_\theta G_n(\theta) - B(\theta;\theta_0)$ and $\theta \mapsto G_n^{sq}(\theta) - C(\theta,\theta_0)$ are continuous on Θ. Thus, using Lemma 2.A.9 together with (2.4.13), (2.4.14), (2.4.15) and (2.4.16) completes the proof of Lemma 2.4.1.(iii).

Finally, Lemma 2.4.1.(iv) follows by an application of Lemma 2.A.10. □

Proof of Lemma 2.4.2. The overall strategy in this proof is to expand the expression on the left-hand side of (2.4.5) in such a manner that all terms either converge to 0 by Lemma 2.A.3, or are equal to 0 by the martingale properties of stochastic integral terms obtained by use of Itô's formula.

By Assumption 2.2.5 and Lemma 2.2.7, the formulae

$$\begin{aligned} g(0,y,x;\theta) &= \tfrac{1}{2}(y-x)^2\partial_y^2 g(0,x,x;\theta) + (y-x)^3 R(y,x;\theta) \\ g^{(1)}(y,x;\theta) &= g^{(1)}(x,x;\theta) + (y-x)R(y,x;\theta) \end{aligned} \tag{2.4.17}$$

may be obtained. Using (2.2.4) and (2.4.17),

$$\begin{aligned} &\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right) \\ &\quad = \mathbb{E}_{\theta_0}\left(\tfrac{1}{2}(X_{t_i^n} - X_{t_{i-1}^n})^2\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right) \\ &\qquad + \mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n})^3 R(X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right) \\ &\qquad + \Delta_n\mathbb{E}_{\theta_0}\left(g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right) \\ &\qquad + \Delta_n\mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n})R(X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right) \\ &\qquad + \Delta^2\mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right). \end{aligned} \tag{2.4.18}$$

Note that

$$\Delta_n g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta_0)\mathbb{E}_{\theta_0}\left(W_{t_i^n} - W_{t_{i-1}^n} \mid \mathcal{F}_{t_{i-1}^n}\right) = 0,$$

and that by repeated use of the Cauchy-Schwarz inequality, Lemma 2.A.4 and Corollary 2.A.5,

$$\begin{aligned} \left|\mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n})^3 R(X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\right| &\leq \Delta_n^2 C(1 + |X_{t_{i-1}^n}|^C) \\ \Delta_n\left|\mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n})R(X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\right| &\leq \Delta_n^2 C(1 + |X_{t_{i-1}^n}|^C) \\ \Delta_n^2\left|\mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\right| &\leq \Delta_n^{5/2} C(1 + |X_{t_{i-1}^n}|^C) \end{aligned}$$

for suitable constants $C > 0$, with

$$\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{[nt]}\Delta_n^{m/2}C(1 + |X_{t_{i-1}^n}|^C) \xrightarrow{\mathcal{P}} 0$$

for $m = 4, 5$ by Lemma 2.A.3. Now, by (2.4.18), it only remains to show that

$$\frac{1}{\sqrt{\Delta_n}}\sum_{i=1}^{[nt]}\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta_0)\mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n})^2(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right) \xrightarrow{\mathcal{P}} 0. \tag{2.4.19}$$

Applying Itô's formula with the function

$$f(y,w) = (y - x_{t^n_{i-1}})^2 (w - w_{t^n_{i-1}})$$

to the process $(X_t, W_t)_{t \geq t^n_{i-1}}$, conditioned on $(X_{t^n_{i-1}}, W_{t^n_{i-1}}) = (x_{t^n_{i-1}}, w_{t^n_{i-1}})$, it follows that

$$\begin{aligned}(X_{t^n_i} - X_{t^n_{i-1}})^2 (W_{t^n_i} - W_{t^n_{i-1}}) \\ &= 2\int_{t^n_{i-1}}^{t^n_i} (X_s - X_{t^n_{i-1}})(W_s - W_{t^n_{i-1}}) a(X_s)\,ds + \int_{t^n_{i-1}}^{t^n_i} (W_s - W_{t^n_{i-1}}) b^2(X_s;\theta_0)\,ds \\ &\quad + 2\int_{t^n_{i-1}}^{t^n_i} (X_s - X_{t^n_{i-1}}) b(X_s;\theta_0)\,ds + 2\int_{t^n_{i-1}}^{t^n_i} (X_s - X_{t^n_{i-1}})(W_s - W_{t^n_{i-1}}) b(X_s;\theta_0)\,dW_s \\ &\quad + \int_{t^n_{i-1}}^{t^n_i} (X_s - X_{t^n_{i-1}})^2\,dW_s\,.\end{aligned} \tag{2.4.20}$$

By the martingale property of the Itô integrals in (2.4.20),

$$\begin{aligned}\mathbb{E}_{\theta_0}&\left((X_{t^n_i} - X_{t^n_{i-1}})^2 (W_{t^n_i} - W_{t^n_{i-1}}) \mid \mathcal{F}_{t^n_{i-1}}\right) \\ &= 2\int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}})(W_s - W_{t^n_{i-1}}) a(X_s) \mid \mathcal{F}_{t^n_{i-1}}\right) ds \\ &\quad + \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((W_s - W_{t^n_{i-1}}) b^2(X_s;\theta_0) \mid \mathcal{F}_{t^n_{i-1}}\right) ds \\ &\quad + 2\int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}}) b(X_s;\theta_0) \mid X_{t^n_{i-1}}\right) ds\,.\end{aligned} \tag{2.4.21}$$

Using the Cauchy-Schwarz inequality, Lemma 2.A.4 and Corollary 2.A.5 again,

$$\left| \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}})(W_s - W_{t^n_{i-1}}) a(X_s) \mid \mathcal{F}_{t^n_{i-1}}\right) ds \right| \leq C\Delta_n^2 (1 + |X_{t^n_{i-1}}|^C)\,,$$

and by Lemma 2.2.6

$$\mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}}) b(X_s;\theta_0) \mid X_{t^n_{i-1}}\right) = (s - t^n_{i-1}) R(s - t^n_{i-1}, X_{t^n_{i-1}};\theta_0)\,,$$

so also

$$\left| \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}}) b(X_s;\theta_0) \mid X_{t^n_{i-1}}\right) ds \right| \leq C\Delta_n^2 (1 + |X_{t^n_{i-1}}|^C)\,.$$

Now

$$\begin{aligned}&\left| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \partial_y^2 g(0, X_{t^n_{i-1}}, X_{t^n_{i-1}};\theta_0) \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}})(W_s - W_{t^n_{i-1}}) a(X_s) \mid \mathcal{F}_{t^n_{i-1}}\right) ds \right| \\ &\quad + \left| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \partial_y^2 g(0, X_{t^n_{i-1}}, X_{t^n_{i-1}};\theta_0) \int_{t^n_{i-1}}^{t^n_i} \mathbb{E}_{\theta_0}\left((X_s - X_{t^n_{i-1}}) b(X_s;\theta_0) \mid X_{t^n_{i-1}}\right) ds \right|\end{aligned}$$

$$\leq \Delta_n^{3/2} C \sum_{i=1}^{[nt]} \left| \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0) \right| (1 + |X_{t_{i-1}^n}|^C)$$
$$\xrightarrow{\mathcal{P}} 0$$

by Lemma 2.A.3, so by (2.4.19) and (2.4.21), it remains to show that

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0) \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((W_s - W_{t_{i-1}^n}) b^2(X_s; \theta_0) \mid \mathcal{F}_{t_{i-1}^n} \right) ds \xrightarrow{\mathcal{P}} 0 .$$

This time, applying Itô's formula with the function

$$f(y, w) = (w - w_{t_{i-1}^n}) b^2(y; \theta_0) ,$$

and making use of the martingale properties of the stochastic integral terms, yields

$$\begin{aligned}
&\int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((W_s - W_{t_{i-1}^n}) b^2(X_s; \theta_0) \mid \mathcal{F}_{t_{i-1}^n} \right) ds \\
&= \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} \mathbb{E}_{\theta_0} \left(a(X_u) \partial_y b^2(X_u; \theta_0)(W_u - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) du\, ds \\
&\quad + \frac{1}{2} \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} \mathbb{E}_{\theta_0} \left(b^2(X_u; \theta_0) \partial_y^2 b^2(X_u; \theta_0)(W_u - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) du\, ds \\
&\quad + \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} \mathbb{E}_{\theta_0} \left(b(X_u; \theta_0) \partial_y b^2(X_u; \theta_0) \mid \mathcal{F}_{t_{i-1}^n} \right) du\, ds .
\end{aligned}$$

Again, by repeated use of the Cauchy-Schwarz inequality and Corollary 2.A.5,

$$\left| \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((W_{t_i^n} - W_{t_{i-1}^n}) b^2(X_s; \theta_0) \mid \mathcal{F}_{t_{i-1}^n} \right) ds \right| \leq C(1 + |X_{t_{i-1}^n}|^C)(\Delta_n^2 + \Delta_n^{5/2}) .$$

Now

$$\begin{aligned}
&\left| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0) \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((W_s - W_{t_{i-1}^n}) b^2(X_s; \theta_0) \mid \mathcal{F}_{t_{i-1}^n} \right) ds \right| \\
&\quad \leq \left(\Delta_n^{3/2} + \Delta_n^2 \right) \sum_{i=1}^{[nt]} \left| \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0) \right| C(1 + |X_{t_{i-1}^n}|^C) \\
&\quad \xrightarrow{\mathcal{P}} 0 ,
\end{aligned}$$

thus completing the proof. □

Proof of Lemma 2.4.4. The aim of this proof is to establish that the conditions of Theorem 2.B.1 (Jacod and Shiryaev, 2003, Theorem IX.7.28) hold, by which the desired result follows directly.

For all $t > 0$,

$$\sup_{s \leq t} \left| \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[ns]} \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \right| \leq \frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \left| \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \right|$$

and since the right-hand side converges to 0 in probability under $\mathbb{P}_{\theta_0}$ by (2.4.1) of Lemma 2.4.1, so does the left-hand side, i.e. Theorem 2.B.1.(i) holds. From (2.4.2) and (2.4.4) of Lemma 2.4.1, it follows that for all $t > 0$,

$$\frac{1}{\Delta_n} \sum_{i=1}^{[nt]} \left(\mathbb{E}_{\theta_0} \left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) - \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right)^2 \right)$$
$$\xrightarrow{\mathcal{P}} \tfrac{1}{2} \int_0^t b^4(X_s; \theta_0) \partial_y^2 g(0, X_s, X_s; \theta_0)^2 \, ds \,,$$

establishing that Theorem 2.B.1.(ii) is satisfied. By Lemma 2.4.2, for all $t > 0$,

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} 0 \,,$$

which corresponds to Theorem 2.B.1.(iii). Finally, by (2.4.3) of Lemma 2.4.1, for all $t > 0$, the Lyapunov condition

$$\frac{1}{\Delta_n^2} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} 0$$

holds, implying the Lindeberg condition of Theorem 2.B.1.(iv). Now, by Theorem 2.B.1, the desired result follows.

It should be noted that the original Theorem IX.7.28 of Jacod and Shiryaev (2003) contains an additional convergence in probability condition. This condition has the same form as Theorem 2.B.1.(iii), but with $W_{t_i^n} - W_{t_{i-1}^n}$ replaced by $N_{t_i^n} - N_{t_{i-1}^n}$, where $\mathbf{N} = (N_t)_{t \geq 0}$ is a placeholder for all bounded martingales on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}_{\theta_0})$, which are orthogonal to $\mathbf{W}$. However, since $(\mathcal{F}_t)_{t \geq 0}$ is generated by U and $\mathbf{W}$, it follows from the martingale representation theorem (Jacod and Shiryaev, 2003, Theorem III.4.33) that every martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \geq 0}, \mathbb{P}_{\theta_0})$ may be written as the sum of a constant term and a stochastic integral with respect to $\mathbf{W}$, and cannot therefore be orthogonal to $\mathbf{W}$. □

Appendix

This section contains a number of technical results utilised in the proofs given in Section 2.4.3.

2.A Auxiliary Results

Lemma 2.A.1. *(Genon-Catalot and Jacod, 1993, Lemma 9) For $i, n \in \mathbb{N}$, let $\mathcal{F}_{n,i} = \mathcal{F}_{t_i^n}$ (with $\mathcal{F}_{n,0} = \mathcal{F}_0$), and let $F_{n,i}$ be an $\mathcal{F}_{n,i}$-measurable, real-valued random variable. If*

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_0}(F_{n,i} \mid \mathcal{F}_{n,i-1}) \xrightarrow{\mathcal{P}} F \quad \text{and} \quad \sum_{i=1}^{n} \mathbb{E}_{\theta_0}(F_{n,i}^2 \mid \mathcal{F}_{n,i-1}) \xrightarrow{\mathcal{P}} 0\,,$$

for some random variable F, then

$$\sum_{i=1}^{n} F_{n,i} \xrightarrow{\mathcal{P}} F\,.$$

◇

Lemma 2.A.1 is taken, without proof, from the paper of Genon-Catalot and Jacod.

Lemma 2.A.2. *Suppose that Assumptions 2.2.4 and 2.2.5 hold. Then, for all $\theta \in \Theta$,*

(i)

$$\begin{aligned}&\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right)\\ &\quad = \tfrac{1}{2}\Delta_n\left(b^2(X_{t_{i-1}^n};\theta_0) - b^2(X_{t_{i-1}^n};\theta)\right)\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,\end{aligned} \tag{2.A.1}$$

(ii)

$$\begin{aligned}&\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right)\\ &\quad = \tfrac{1}{2}\Delta_n\left(b^2(X_{t_{i-1}^n};\theta_0) - b^2(X_{t_{i-1}^n};\theta)\right)\partial_y^2 \partial_\theta g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\\ &\qquad - \tfrac{1}{2}\Delta_n \partial_\theta b^2(X_{t_{i-1}^n};\theta)\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,\end{aligned} \tag{2.A.2}$$

(iii)

$$\begin{aligned}&\mathbb{E}_{\theta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right)\\ &\quad = \tfrac{1}{2}\Delta_n^2\left(b^4(X_{t_{i-1}^n};\theta_0) + \tfrac{1}{2}\left(b^2(X_{t_{i-1}^n};\theta_0) - b^2(X_{t_{i-1}^n};\theta)\right)^2\right)\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^2\\ &\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,\end{aligned} \tag{2.A.3}$$

(iv)

$$\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)^2 \mid X_{t_{i-1}^n}\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,, \tag{2.A.4}$$

(v)

$$\mathbb{E}_{\theta_0}\left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^4 R(\Delta_n, X_{t_{i-1}^n};\theta). \tag{2.A.5}$$

◇

Proof of Lemma 2.A.2. For sake of completeness, a proof of all five formulae is given, although (2.A.1), (2.A.2) and (2.A.3) are already implicitly given in the proofs of (Sørensen, 2010, Lemmas 3.2 & 3.4). Note first that using (2.2.5),

$$\begin{aligned}
\mathcal{L}_{\theta_0}(g(0,\theta))(x,x) &= \tfrac{1}{2}b^2(x;\theta_0)\partial_y^2 g(0,x,x;\theta)\\
\mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta))(x,x) &= \tfrac{1}{2}b^2(x;\theta_0)\partial_y^2\partial_\theta g(0,x,x;\theta)\\
\mathcal{L}_{\theta_0}^2(g^2(0;\theta))(x,x) &= \tfrac{3}{2}b^4(x;\theta_0)\partial_y^2 g(0,x,x;\theta)^2\\
\mathcal{L}_{\theta_0}(g(0,\theta)g^{(1)}(\theta))(x,x) &= -\tfrac{1}{4}b^2(x;\theta)b^2(x;\theta_0)\partial_y^2 g(0,x,x;\theta)^2
\end{aligned}$$

and

$$\begin{aligned}
\mathcal{L}_{\theta_0}(g^2(0;\theta))(x,x) &= 0\\
\mathcal{L}_{\theta_0}^i(g^4(0;\theta))(x,x) &= 0\,, \quad i=1,2,3\\
\mathcal{L}_{\theta_0}^i(g^3(0,\theta)g^{(1)}(\theta))(x,x) &= 0\,, \quad i=1,2\\
\mathcal{L}_{\theta_0}(g^2(0,\theta)g^{(1)}(\theta)^2)(x,x) &= 0\\
\mathcal{L}_{\theta_0}(g^3(0,\theta)g^{(2)}(\theta))(x,x) &= 0\\
\mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta)^2)(x,x) &= 0\,.
\end{aligned}$$

The verification of these formulae may be simplified by using e.g. the Leibniz formula for the n'th derivative of a product, together with the results of Lemma 2.2.7 and Assumption 2.2.5.(ii), to see that many of the partial derivatives which appear during the process are zero when evaluated in $y = x$. These results, as well as Lemmas 2.2.6 and 2.2.7, and (2.A.8) are used without reference in the following.

First, see that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad= \mathbb{E}_{\theta_0}\left(g(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) + \Delta_n\mathbb{E}_{\theta_0}\left(g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad+ \Delta_n^2\mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad= g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n\mathcal{L}_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\\
&\qquad+ \Delta_n\left(g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\right) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad= \tfrac{1}{2}\Delta_n\left(b^2(X_{t_{i-1}^n};\theta_0) - b^2(X_{t_{i-1}^n};\theta)\right)\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}$$

which proves Lemma 2.A.2.(i). Now, using that

$$\partial_\theta\mathcal{L}_\theta(g(0,\theta))(x,x) = \mathcal{L}_\theta(\partial_\theta g(0,\theta))(x,x) + \tfrac{1}{2}\partial_\theta\, b^2(x;\theta)\partial_y^2 g(0,x,x;\theta)\,,$$

it follows that

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(\partial_\theta g(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) + \Delta_n \mathbb{E}_{\theta_0}\left(\partial_\theta g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \partial_\theta g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n \mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\qquad + \Delta_n\left(\partial_\theta g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\right) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \Delta_n\left(\mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - \partial_\theta \mathcal{L}_\theta(g(0,\theta))(X_{t_i^n}, X_{t_{i-1}^n})\right) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \tfrac{1}{2}\Delta_n\left(b^2(X_{t_{i-1}^n};\theta_0) - b^2(X_{t_{i-1}^n};\theta)\right)\partial_y^2 \partial_\theta g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\\
&\qquad - \tfrac{1}{2}\Delta_n \partial_\theta b^2(X_{t_{i-1}^n};\theta)\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}
$$

thus proving Lemma 2.A.2.(ii). Furthermore,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(\partial_\theta g(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right)\\
&\qquad + 2\Delta_n \mathbb{E}_{\theta_0}\left(\partial_\theta g(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)\partial_\theta g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \partial_\theta g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^2 + \Delta_n \mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta)^2)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\qquad + 2\Delta_n\left(\partial_\theta g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\partial_\theta g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\right)\\
&\quad = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}
$$

proving Lemma 2.A.2.(iv). Similarly,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(g^2(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) + 2\Delta_n \mathbb{E}_{\theta_0}\left(g(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 + g(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)g^{(2)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^3 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = g^2(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n \mathcal{L}_{\theta_0}(g^2(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \tfrac{1}{2}\Delta_n^2 \mathcal{L}_{\theta_0}^2(g^2(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\\
&\qquad + 2\Delta_n\left(g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n \mathcal{L}_{\theta_0}(g(0;\theta)g^{(1)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)\\
&\qquad + \Delta_n^2\left(g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^2 + g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)g^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\right) + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \Delta_n^2\left(\tfrac{1}{2}\mathcal{L}_{\theta_0}^2(g^2(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + 2\mathcal{L}_{\theta_0}(g(0;\theta)g^{(1)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)\\
&\qquad + \Delta_n^2\left(\mathcal{L}_\theta(g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)^2 + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \tfrac{1}{2}\Delta_n^2\left(b^4(X_{t_{i-1}^n};\theta_0) + \tfrac{1}{2}\left(b^2(X_{t_{i-1}^n};\theta_0) - b^2(X_{t_{i-1}^n};\theta)\right)^2\right)\partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^2\\
&\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}
$$

and

$$
\mathbb{E}_{\theta_0}\left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)
$$

$$
\begin{aligned}
&= \mathbb{E}_{\theta_0}\left(g^4(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + 4\Delta_n \mathbb{E}_{\theta_0}\left(g^3(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + 6\Delta_n^2 \mathbb{E}_{\theta_0}\left(g^2(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta)^2 \mid X_{t_{i-1}^n}\right) \\
&\quad + 2\Delta_n^2 \mathbb{E}_{\theta_0}\left(g^3(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(2)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + 4\Delta_n^3 \mathbb{E}_{\theta_0}\left(g(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta)^3 \mid X_{t_{i-1}^n}\right) \\
&\quad + 6\Delta_n^3 \mathbb{E}_{\theta_0}\left(g^2(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(2)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + \tfrac{2}{3}\Delta_n^3 \mathbb{E}_{\theta_0}\left(g^3(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g^{(3)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + \Delta_n^4 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&= g^4(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n \mathcal{L}_{\theta_0}(g^4(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \tfrac{1}{2}\Delta_n^2 \mathcal{L}_{\theta_0}^2(g^4(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\quad + \tfrac{1}{6}\Delta_n^3 \mathcal{L}_{\theta_0}^3(g^4(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + 4\Delta_n g^3(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\quad + 4\Delta_n^2 \mathcal{L}_{\theta_0}(g^3(0;\theta) g^{(1)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + 2\Delta_n^3 \mathcal{L}_{\theta_0}^2(g^3(0;\theta) g^{(1)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\quad + 6\Delta_n^2 g^2(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta)^2 + 6\Delta_n^3 \mathcal{L}_{\theta_0}(g^2(0;\theta) g^{(1)}(\theta)^2)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\quad + 2\Delta_n^2 g^3(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + 2\Delta_n^3 \mathcal{L}_{\theta_0}(g^3(0;\theta) g^{(2)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\quad + 4\Delta_n^3 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta)^3 \\
&\quad + 6\Delta_n^3 g^2(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\quad + \tfrac{2}{3}\Delta_n^3 g^3(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g^{(3)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\quad + \Delta_n^4 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&= \Delta_n^4 R(\Delta_n, X_{t_{i-1}^n}; \theta)\,,
\end{aligned}
$$

which prove Lemma 2.A.2.(iii) and (v) as well. □

Lemma 2.A.3. *Let $x \mapsto f(x)$ be a continuous, real-valued function, and let $t > 0$ be given. Then*

$$
\Delta_n \sum_{i=1}^{[nt]} f(X_{t_{i-1}^n}) \xrightarrow{\mathcal{P}} \int_0^t f(X_s)\, ds\,.
$$

◇

Lemma 2.A.3 follows easily by the convergence of Riemann sums, and is presented without proof.

Lemma 2.A.4. *Suppose that Assumption 2.2.4 holds, and let $m \geq 2$. Then, there exists a constant $C_m > 0$, such that for $0 \leq t \leq t + \Delta \leq 1$,*

$$
\mathbb{E}_{\theta_0}\left(|X_{t+\Delta} - X_t|^m \mid X_t\right) \leq C_m \Delta^{m/2}\left(1 + |X_t|^m\right)\,. \tag{2.A.6}
$$

◇

Corollary 2.A.5. *Suppose that Assumption 2.2.4 holds. Let a compact, convex set $K \subseteq \Theta$ be given, and suppose that $f(y, x; \theta)$ is of polynomial growth in x and y, uniformly for θ in*

K. Then, there exist constants $C_K > 0$ such that for $0 \leq t \leq t + \Delta \leq 1$,

$$\mathbb{E}_{\theta_0}\left(|f(X_{t+\Delta}, X_t, \theta)| \mid X_t\right) \leq C_K\left(1 + |X_t|^{C_K}\right)$$

for all $\theta \in K$. ◇

Lemma 2.A.4 and Corollary 2.A.5, correspond to Lemma 6 of Kessler (1997), adapted to the present assumptions.[5] The corollary is a simple consequence of the lemma. For use in the following, observe that for any $\theta \in \Theta$, there exists a constant $C_\theta > 0$ such that

$$\Delta_n \sum_{i=1}^{[nt]} \left|R_\theta(\Delta_n, X_{t_{i-1}^n})\right| \leq C_\theta \Delta_n \sum_{i=1}^{[nt]} \left(1 + |X_{t_{i-1}^n}|^{C_\theta}\right),$$

so it follows from Lemma 2.A.3 that for any deterministic, real-valued sequence $(\delta_n)_{n\in\mathbb{N}}$ with $\delta_n \to 0$ as $n \to \infty$,

$$\delta_n \Delta_n \sum_{i=1}^{[nt]} \left|R_\theta(\Delta_n, X_{t_{i-1}^n})\right| \xrightarrow{\mathcal{P}} 0\,. \tag{2.A.7}$$

In particular, (2.A.7) holds for $R(\Delta_n, X_{t_{i-1}^n}; \theta)$. Also, note that by Corollary 2.A.5, it holds that under Assumption 2.2.4,

$$\mathbb{E}_{\theta_0}\left(R\left(\Delta, X_{t+\Delta}, X_t; \theta\right) \mid X_t\right) = R(\Delta, X_t; \theta)\,. \tag{2.A.8}$$

Lemma 2.A.6. *Suppose that Assumption 2.2.4 holds, and that the function $f(t, y, x; \theta)$ satisfies that*

$$f(t, y, x; \theta) \in C^{pol}_{1,2,1}([0,1] \times \mathcal{X}^2 \times \Theta) \quad \textit{with} \quad f(0, x, x; \theta) = 0$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Then, for all $\theta \in \Theta$,

$$f(t-s, X_t, X_s; \theta) = \int_s^t f_1\left(u-s, X_u, X_s; \theta\right) du + \int_s^t f_2\left(u-s, X_u, X_s; \theta\right) dW_u \tag{2.A.9}$$

under $\mathbb{P}_{\theta_0}$, where f_1 and f_2 are given by

$$\begin{aligned} f_1(t, y, x; \theta) &= \partial_t f(t, y, x; \theta) + a(y)\partial_y f(t, y, x; \theta) + \tfrac{1}{2} b^2(y; \theta_0)\partial_y^2 f(t, y, x; \theta) \\ f_2(t, y, x; \theta) &= b(y; \theta_0)\partial_y f(t, y, x; \theta)\,. \end{aligned}$$

Furthermore, let $m \in \mathbb{N}$ be given, and let $Dk(\,\cdot\,; \theta, \theta') = k(\,\cdot\,; \theta) - f(\,\cdot\,; \theta')$. Then, there exist constants $C_m > 0$ such that

$$\begin{aligned} &\mathbb{E}_{\theta_0}\left(\left|Df(t-s, X_t, X_s; \theta, \theta')\right|^{2m}\right) \\ &\quad\leq C_m(t-s)^{2m-1} \int_s^t \mathbb{E}_{\theta_0}\left(\left|Df_1(u-s, X_u, X_s; \theta, \theta')\right|^{2m}\right) du \\ &\qquad + C_m(t-s)^{m-1} \int_s^t \mathbb{E}_{\theta_0}\left(\left|Df_2(u-s, X_u, X_s; \theta, \theta')\right|^{2m}\right) du \end{aligned} \tag{2.A.10}$$

[5] Section 3.A.3 contains a proof of Lemma 3.A.23, an inequality for jump-diffusions which resembles (2.A.6). As this proof is essentially an extended version of the proof given by Flachs (2011), of the inequality (2.A.6) for (ergodic) continuous diffusions, it is easily modified to prove Lemma 2.A.4.

for $0 \leq s < t \leq 1$ *and* $\theta, \theta' \in \Theta$*. Also, for each compact, convex set* $K \subseteq \Theta$*, there exists a constant* $C_{m,K} > 0$ *such that*

$$\mathbb{E}_{\theta_0}\left(|Df_j(t-s, X_t, X_s; \theta, \theta')|^{2m}\right) \leq C_{m,K}|\theta - \theta'|^{2m}$$

for $j = 1, 2$, $0 \leq s < t \leq 1$ *and all* $\theta, \theta' \in K$. ◇

Proof of Lemma 2.A.6. A simple application of Itô's formula (when conditioning on $X_s = x_s$) yields (2.A.9).

By Jensen's inequality, it holds that for any $k \in \mathbb{N}$,

$$\mathbb{E}_{\theta_0}\left(\left|\int_s^t Df_j(u-s, X_u, X_s; \theta, \theta')^j\, du\right|^k\right) \leq (t-s)^{k-1}\int_s^t \mathbb{E}_{\theta_0}\left(\left|Df_j(u-s, X_u, X_s; \theta, \theta')\right|^{jk}\right) du \tag{2.A.11}$$

for $j = 1, 2$, and by the martingale properties of the second term in (2.A.9), the Burkholder-Davis-Gundy inequality may be used to show that

$$\mathbb{E}_{\theta_0}\left(\left|\int_s^t Df_2(u-s, X_u, X_s; \theta, \theta')\, dW_u\right|^{2m}\right) \leq C_m \mathbb{E}_{\theta_0}\left(\left|\int_s^t Df_2(u-s, X_u, X_s; \theta, \theta')^2\, du\right|^m\right). \tag{2.A.12}$$

Now, (2.A.9), (2.A.11) and (2.A.12) may be combined to show (2.A.10). The last result of the lemma follows by a simple application of the mean value theorem. □

Lemma 2.A.7. *Suppose that Assumption 2.2.4 holds, and let* $K \subseteq \Theta$ *be compact and convex. Assume that*

$$f(t, y, x; \theta) \in C^{pol}_{1,2,1}([0,1] \times \mathcal{X}^2 \times \Theta) \quad \textit{with} \quad f(0, x, x; \theta) = 0 \tag{2.A.13}$$

for all $x \in \mathcal{X}$ *and* $\theta \in \Theta$*, and define*

$$F_n(\theta) = \sum_{i=1}^n f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta).$$

Then, for each $m \in \mathbb{N}$*, there exists a constant* $C_{m,K} > 0$*, such that*

$$\mathbb{E}_{\theta_0}\left|F_n(\theta) - F_n(\theta')\right|^{2m} \leq C_{m,K}\,|\theta - \theta'|^{2m}$$

for all $\theta, \theta' \in K$ *and* $n \in \mathbb{N}$*. Suppose, furthermore, that the functions*

$$\begin{aligned} h_1(t, y, x; \theta) &= \partial_t f(t, y, x; \theta) + a(y)\partial_y f(t, y, x; \theta) + \tfrac{1}{2}b^2(y; \theta_0)\partial_y^2 f(t, y, x; \theta) \\ h_2(t, y, x; \theta) &= b(y; \theta_0)\partial_y f(t, y, x; \theta) \\ h_{j2}(t, y, x; \theta) &= b(y; \theta_0)\partial_y h_j(t, y, x, \theta) \end{aligned}$$

also satisfy (2.A.13) for $j = 1, 2$*, and define*

$$\widetilde{F}_n(\theta) = \frac{1}{\Delta_n}\sum_{i=1}^n f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta).$$

Then, for each $m \in \mathbb{N}$, there exists a constant $C_{m,K} > 0$, such that

$$\mathbb{E}_{\theta_0} \left| \widetilde{F}_n(\theta) - \widetilde{F}_n(\theta') \right|^{2m} \leq C_{m,K} \, |\theta - \theta'|^{2m}$$

for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$. ◇

Proof of Lemma 2.A.7. This proof is a rewriting of the proof of Sørensen (2010, Lemma 5.5). For use in the following, define, in addition to h_1, h_2 and h_{j2}, the functions

$$\begin{aligned}
h_{j1}(t, y, x; \theta) &= \partial_t h_j(t, y, x; \theta) + a(y)\partial_y h_j(t, y, x; \theta) + \tfrac{1}{2} b^2(y; \theta_0)\partial_y^2 h_j(t, y, x; \theta) \\
h_{j21}(t, y, x; \theta) &= \partial_t h_{j2}(t, y, x; \theta) + a(y)\partial_y h_{j2}(t, y, x; \theta) + \tfrac{1}{2} b^2(y; \theta_0)\partial_y^2 h_{j2}(t, y, x; \theta) \\
h_{j22}(t, y, x; \theta) &= b(y; \theta_0)\partial_y h_{j2}(t, y, x; \theta)
\end{aligned}$$

for $j = 1, 2$, and, for ease of notation, let

$$H_j^{n,i}(u; \theta, \theta') = Dh_j(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}; \theta, \theta')$$

for $j \in \{1, 2, 11, 12, 21, 22, 121, 122, 221, 222\}$, where $Dk(\,\cdot\,; \theta, \theta') = k(\,\cdot\,; \theta) - f(\,\cdot\,; \theta')$. Recall also that $\Delta_n = 1/n$.

First, by the martingale properties of

$$\Delta_n \sum_{i=1}^n \int_0^r \mathbf{1}_{(t_{i-1}^n, t_i^n]}(u) H_2^{n,i}(u; \theta, \theta') \, dW_u \,,$$

the Burkholder-Davis-Gundy inequality is used to establish the existence of a constant $C_m > 0$ such that

$$\mathbb{E}_{\theta_0}\left(\left| \Delta_n \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} H_2^{n,i}(u; \theta, \theta') \, dW_u \right|^{2m} \right) \leq C_m \mathbb{E}_{\theta_0}\left(\left| \Delta_n^2 \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} H_2^{n,i}(u; \theta, \theta')^2 \, du \right|^{m} \right).$$

Now, using also Jensen's inequality and Lemma 2.A.6,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left| \Delta_n \sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta') \right|^{2m} \right) \\
&\leq C_m \mathbb{E}_{\theta_0}\left(\left| \Delta_n \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} H_1^{n,i}(u; \theta, \theta') \, du \right|^{2m} \right) + C_m \mathbb{E}_{\theta_0}\left(\left| \Delta_n \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} H_2^{n,i}(u; \theta, \theta') \, dW_u \right|^{2m} \right) \\
&\leq C_m \Delta_n \sum_{i=1}^n \mathbb{E}_{\theta_0}\left(\left| \int_{t_{i-1}^n}^{t_i^n} H_1^{n,i}(u; \theta, \theta') \, du \right|^{2m} \right) + C_m \mathbb{E}_{\theta_0}\left(\left| \Delta_n^2 \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} H_2^{n,i}(u; \theta, \theta')^2 \, du \right|^{m} \right) \\
&\leq C_m \Delta_n^{2m+1} \sum_{i=1}^n \left(\mathbb{E}_{\theta_0}\left(\left| \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} H_1^{n,i}(u; \theta, \theta') \, du \right|^{2m} \right) + \mathbb{E}_{\theta_0}\left(\left| \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} H_2^{n,i}(u; \theta, \theta')^2 \, du \right|^{m} \right) \right) \\
&\leq C_m \Delta_n^{2m} \sum_{i=1}^n \left(\int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(|H_1^{n,i}(u; \theta, \theta')|^{2m} \right) du + \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(|H_2^{n,i}(u; \theta, \theta')|^{2m} \right) du \right) \qquad (2.A.14) \\
&\leq C_{m,K} |\theta - \theta'|^{2m} \Delta_n^{2m} \,,
\end{aligned}$$

thus

$$\mathbb{E}_{\theta_0}\left(|DF_n(\theta,\theta')|^{2m}\right) = \Delta_n^{-2m}\mathbb{E}_{\theta_0}\left(\left|\Delta_n \sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta,\theta')\right|^{2m}\right) \leq C_{m,K}|\theta-\theta'|^{2m}$$

for all $\theta,\theta' \in K$ and $n \in \mathbb{N}$. In the case where also h_j and h_{j2} satisfy (2.A.13) for all $x \in \mathcal{X}$, $\theta \in \Theta$ and $j = 1, 2$, use Lemma 2.A.6 to write

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(|H_1^{n,i}(u;\theta,\theta')|^{2m}\right) \\
&\quad\leq C_m(u-t_{i-1}^n)^{2m-1}\int_{t_{i-1}^n}^u \mathbb{E}_{\theta_0}\left(|H_{11}^{n,i}(v;\theta,\theta')|^{2m}\right)dv \\
&\qquad + C_m(u-t_{i-1}^n)^{m-1}\int_{t_{i-1}^n}^u \mathbb{E}_{\theta_0}\left(|H_{12}^{n,i}(v;\theta,\theta')|^{2m}\right)dv \\
&\quad\leq C_m(u-t_{i-1}^n)^{2m-1}\int_{t_{i-1}^n}^u \mathbb{E}_{\theta_0}\left(|H_{11}^{n,i}(v;\theta,\theta')|^{2m}\right)dv \\
&\qquad + C_m(u-t_{i-1}^n)^{m-1}\int_{t_{i-1}^n}^u \left((v-t_{i-1}^n)^{2m-1}\int_{t_{i-1}^n}^v \mathbb{E}_{\theta_0}\left(\left|H_{121}^{n,i}(w;\theta,\theta')\right|^{2m}\right)dw\right)dv \\
&\qquad + C_m(u-t_{i-1}^n)^{m-1}\int_{t_{i-1}^n}^u \left((v-t_{i-1}^n)^{m-1}\int_{t_{i-1}^n}^v \mathbb{E}_{\theta_0}\left(\left|H_{122}^{n,i}(w;\theta,\theta')\right|^{2m}\right)dw\right)dv \\
&\quad\leq C_{m,K}|\theta-\theta'|^{2m}\left((u-t_{i-1}^n)^{2m}+(u-t_{i-1}^n)^{3m}\right),
\end{aligned}$$

and similarly obtain

$$\mathbb{E}_{\theta_0}\left(|H_2^{n,i}(u;\theta,\theta')|^{2m}\right) \leq C_{m,K}|\theta-\theta'|^{2m}\left((u-t_{i-1}^n)^{2m}+(u-t_{i-1}^n)^{3m}\right).$$

Now, inserting into (2.A.14),

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\Delta_n \sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta,\theta')\right|^{2m}\right) \\
&\quad\leq C_{m,K}\Delta_n^{2m}\sum_{i=1}^n\left(\int_{t_{i-1}^n}^{t_i^n}\mathbb{E}_{\theta_0}\left(|H_1^{n,i}(u;\theta,\theta')|^{2m}\right)du + \int_{t_{i-1}^n}^{t_i^n}\mathbb{E}_{\theta_0}\left(|H_2^{n,i}(u;\theta,\theta')|^{2m}\right)du\right) \\
&\quad\leq C_{m,K}|\theta-\theta'|^{2m}\Delta_n^{2m}\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\left((u-t_{i-1}^n)^{2m}+(u-t_{i-1}^n)^{3m}\right)du \\
&\quad\leq C_{m,K}|\theta-\theta'|^{2m}\left(\Delta_n^{4m}+\Delta_n^{5m}\right),
\end{aligned}$$

and, ultimately,

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(|D\tilde{F}_n(\theta,\theta')|^{2m}\right) &= \mathbb{E}_{\theta_0}\left(\left|\Delta_n^{-1}\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta,\theta')\right|^{2m}\right) \\
&= \Delta_n^{-4m}\mathbb{E}_{\theta_0}\left(\left|\Delta_n\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta,\theta')\right|^{2m}\right) \\
&\leq C_{m,K}|\theta-\theta'|^{2m}(1+\Delta_n) \\
&\leq C_{m,K}|\theta-\theta'|^{2m}.
\end{aligned}$$

□

Lemma 2.A.8. *Suppose that Assumption 2.2.4 is satisfied. Let* $f \in C^{pol}_{0,1}(\mathcal{X} \times \Theta)$*. Define*

$$F(\theta) = \int_0^1 f(X_s; \theta)\, ds$$

and let $K \subseteq \Theta$ *be compact and convex. Then, for each* $m \in \mathbb{N}$*, there exists a constant* $C_{m,K} > 0$ *such that for all* $\theta, \theta' \in K$*,*

$$\mathbb{E}_{\theta_0} |F(\theta) - F(\theta')|^{2m} \leq C_{m,K}\ |\theta - \theta'|^{2m} .$$

◇

Lemma 2.A.8 follows from a simple application of the mean value theorem and is presented without proof.

Lemma 2.A.9. *Let* $K \subseteq \Theta$ *be compact and convex. Suppose that* $\mathbf{H}_n = (H_n(\theta))_{\theta \in K}$ *defines a sequence* $(\mathbf{H}_n)_{n \in \mathbb{N}}$ *of continuous, real-valued stochastic processes such that for all* $\theta \in K$*,*

$$H_n(\theta) \xrightarrow{\mathcal{P}} 0$$

for fixed θ*. Furthermore, assume that for some* $m \in \mathbb{N}$*, there exists a constant* $C_{m,K} > 0$ *such that for all* $\theta, \theta' \in K$ *and* $n \in \mathbb{N}$*,*

$$\mathbb{E}_{\theta_0} \left| H_n(\theta) - H_n(\theta') \right|^{2m} \leq C_{m,K} |\theta - \theta'|^{2m} . \tag{2.A.15}$$

Then,

$$\sup_{\theta \in K} |H_n(\theta)| \xrightarrow{\mathcal{P}} 0 .$$

◇

Proof of Lemma 2.A.9*.* $(H_n(\theta))_{n \in \mathbb{N}}$ is tight in $\mathbb{R}$ for all $\theta \in K$, so, using (2.A.15), it follows from Kallenberg (1997, Corollary 14.9 & Theorem 14.3) that the sequence of processes $(\mathbf{H}_n)_{n \in \mathbb{N}}$ is tight in $C(K, \mathbb{R})$, the space of continuous (and bounded) real-valued functions on K, and thus relatively compact in distribution. Also, for all $d \in \mathbb{N}$ and $(\theta_1, \ldots, \theta_d) \in K^d$,

$$\begin{pmatrix} H_n(\theta_1) \\ \vdots \\ H_n(\theta_d) \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

so by Kallenberg (1997, Lemma 14.2), $\mathbf{H}_n \xrightarrow{\mathcal{D}} 0$ in $C(K, \mathbb{R})$ equipped with the uniform metric. Finally, by the continuous mapping theorem,

$$\sup_{\theta \in K} |H_n(\theta)| \xrightarrow{\mathcal{D}} 0 ,$$

and the desired result follows. □

Lemma 2.A.10. *Let $(H_n(\theta))_{\theta\in\Theta}$, $n \in \mathbb{N}$, and $(H(\theta))_{\theta\in\Theta}$ be real-valued, stochastic processes. Suppose that $(H(\theta))_{\theta\in\Theta}$ is continuous, that*

$$\sup_{\theta\in K} |H_n(\theta) - H(\theta)| \xrightarrow{\mathcal{P}} 0$$

for all compact, convex subsets $K \subseteq \Theta$, and that $\hat{\theta}_n$ is a consistent estimator of θ_0. Then

$$H_n(\hat{\theta}_n) \xrightarrow{\mathcal{P}} H(\theta_0)\,.$$

◇

Proof of Lemma 2.A.10. The objective is to show that for all $\delta, \varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}_{\theta_0}(|H_n(\hat{\theta}_n) - H(\theta_0)| \leq \varepsilon) > 1 - \delta \quad \text{for} \quad n \geq n_0\,. \tag{2.A.16}$$

Choose $\delta, \varepsilon > 0$. Since $\theta \mapsto H(\theta)$ is continuous, there exists $\eta > 0$ such that

$$|\theta - \theta_0| \leq \eta \quad \Rightarrow \quad |H(\theta) - H(\theta_0)| \leq \tfrac{\varepsilon}{2}\,.$$

Let $K = \{\theta \in \Theta \,:\, |\theta - \theta_0| \leq \eta\}$. By assumption, there exist $n_1, n_2 \in \mathbb{N}$ such that

$$\mathbb{P}_{\theta_0}\left(|\hat{\theta}_n - \theta_0| \leq \eta\right) > 1 - \tfrac{\delta}{2} \quad \text{for} \quad n \geq n_1$$

and

$$\mathbb{P}_{\theta_0}\left(\sup_{\theta\in K} |H_n(\theta) - H(\theta)| \leq \tfrac{\varepsilon}{2}\right) > 1 - \tfrac{\delta}{2} \quad \text{for} \quad n \geq n_2\,.$$

Now, let $n_0 = \max\{n_1, n_2\}$, and to conclude (2.A.16), use that on the set

$$\left(|\hat{\theta}_n - \theta_0| \leq \eta\right) \cap \left(\sup_{\theta\in K} |H_n(\theta) - H(\theta)| \leq \tfrac{\varepsilon}{2}\right)$$

it holds that

$$|H_n(\hat{\theta}_n) - H(\theta_0)| \leq \sup_{\theta\in K} |H_n(\theta) - H(\theta)| + |H(\hat{\theta}_n) - H(\theta_0)| \leq \varepsilon\,.$$ □

2.B Theorems from the Literature

In this section, some results from the literature, important to the proof of Theorem 2.3.2, are summarised in a greatly simplified form, tailored specifically to the approximate martingale estimating function-setup of the current paper. Section 2.B.1 contains a version of Theorem IX.7.28 of Jacod and Shiryaev (2003), while Section 2.B.2 contains selected results of Jacod and Sørensen (2012) and Sørensen (2012, Section 1.10).

2.B.1 Stable Limit Theorem

Theorem 2.B.1 below is a simplified version of Theorem IX.7.28 of Jacod and Shiryaev (2003).

Theorem 2.B.1. *For $i, n \in \mathbb{N}$, set $\Delta W_{n,i} = W_{t_i^n} - W_{t_{i-1}^n}$ and $\mathcal{F}_{n,i} = \mathcal{F}_{t_i^n}$ (with $\mathcal{F}_{n,0} = \mathcal{F}_0$), and let $F_{n,i}$ be a square-integrable, $\mathcal{F}_{n,i}$-measurable, real-valued random variable. Let $(C_t)_{t\geq 0}$ be a continuous, $(\mathcal{F}_t)_{t\geq 0}$-adapted, real-valued process of the form*

$$C_t = \int_0^t c_s^2 \, ds \, .$$

Suppose that for all $t > 0$, the following holds.

(i)

$$\sup_{s\leq t} \Big| \sum_{i=1}^{[ns]} \mathbb{E}_{\theta_0} \left(F_{n,i} \mid \mathcal{F}_{n,i-1} \right) \Big| \xrightarrow{\mathcal{P}} 0 \, .$$

(ii)

$$\sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(F_{n,i}^2 \mid \mathcal{F}_{n,i-1} \right) - \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(F_{n,i} \mid \mathcal{F}_{n,i-1} \right)^2 \xrightarrow{\mathcal{P}} C_t \, .$$

(iii)

$$\sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(F_{n,i} \Delta W_{n,i} \mid \mathcal{F}_{n,i-1} \right) \xrightarrow{\mathcal{P}} 0 \, .$$

(iv) For all $\varepsilon > 0$,

$$\sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(F_{n,i}^2 \mathbf{1}(|F_{n,i}| > \varepsilon) \mid \mathcal{F}_{n,i-1} \right) \xrightarrow{\mathcal{P}} 0 \, .$$

Put

$$Y_{n,t} = \sum_{i=1}^{[nt]} F_{n,i} \, .$$

Then, the processes $\mathbf{Y}_n = (Y_{n,t})_{t\geq 0}$ converge stably in distribution under $\mathbb{P}_{\theta_0}$ to the process $\mathbf{Y} = (Y_t)_{t\geq 0}$ given by

$$Y_t = \int_0^t c_s \, dB_s \, .$$

$\mathbf{B} = (B_s)_{s\geq 0}$ is a standard Wiener process, which is independent of U and $\mathbf{W}$, and defined on a filtered extension $(\Omega', \mathcal{F}', (\mathcal{F}'_t)_{t\geq 0}, P'_{\theta_0})$ of $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, P_{\theta_0})$. ⋄

Note that the original theorem (Jacod and Shiryaev, 2003, Theorem IX.7.28) contains an additional convergence in probability condition, which becomes superfluous in the current setup. See the end of the proof of Lemma 2.4.4 for more details.

2.B.2 Asymptotic Results for Estimating Functions

This section briefly summarises Theorems 1.58, 1.59 and 1.60 and some additional comments from Sørensen (2012), and Lemma 2.14 of Jacod and Sørensen (2012), adapted to the setup of the current paper. Proofs of these results are given by Jacod and Sørensen (2012).

In the following, let $G_n(\theta)$ be an approximate martingale estimating function as given in Definition 2.2.2, with associated G_n-estimators defined in Definition 2.2.3.

Theorem 2.B.2. *Sørensen (2012, Theorem 1.58) Suppose that there exist a compact, convex set $K \subseteq \Theta$ with $\theta_0 \in \text{int}\, K$, and a (possibly random) real-valued function $\theta \mapsto B(\theta;\theta_0)$ on K, such that*

(i) $G_n(\theta_0) \xrightarrow{\mathcal{P}} 0.$

(ii) The function $\theta \mapsto G_n(\theta)$ is continuously differentiable on K for all $n \in \mathbb{N}$, with

$$\sup_{\theta\in K} |\partial_\theta G_n(\theta) - B(\theta;\theta_0)| \xrightarrow{\mathcal{P}} 0\,.$$

(iii) $B(\theta_0;\theta_0)$ is non-singular (with probability one under $\mathbb{P}_{\theta_0}$).

Then, there exists a consistent G_n-estimator $\hat{\theta}_n$, which is eventually unique in the sense that for any other consistent G_n-estimator $\bar{\theta}_n$, $\mathbb{P}_{\theta_0}(\hat{\theta}_n \neq \bar{\theta}_n) \to 0$ as $n \to \infty$. ◇

By Sørensen (2012, p. 87), under the conditions of Theorem 2.B.2, the mapping $\theta \mapsto B(\theta;\theta_0)$ is continuous on K (up to a $\mathbb{P}_{\theta_0}$-null set, if $B(\theta;\theta_0)$ is random). Also, there exists a unique, continuously differentiable real-valued function $\theta \mapsto A(\theta;\theta_0)$ (still, up to a $\mathbb{P}_{\theta_0}$-null set), satisfying that $A(\theta_0;\theta_0) = 0$, $\theta \mapsto \partial_\theta A(\theta;\theta_0) = B(\theta;\theta_0)$ for all $\theta \in K$ and

$$\sup_{\theta\in K} |G_n(\theta) - A(\theta;\theta_0)| \xrightarrow{\mathcal{P}} 0\,.$$

Theorem 2.B.3. *Sørensen (2012, Theorem 1.59) Suppose that the conditions of Theorem 2.B.2 are satisfied, and that the aforementioned function $A(\theta;\theta_0)$ satisfies that for all $\varepsilon > 0$,*

$$\mathbb{P}_{\theta_0}\left(\inf_{K\setminus\bar{B}_\varepsilon(\theta_0)} |A(\theta;\theta_0)| > 0\right) = 1\,, \tag{2.B.1}$$

where $\bar{B}_\varepsilon(\theta_0)$ denotes the closed ball in $\mathbb{R}$, with radius ε and centre θ_0. Then, for any G_n-estimator $\tilde{\theta}_n$, it holds that for all $\varepsilon > 0$,

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K\setminus\bar{B}_\varepsilon(\theta_0)\right) \to 0$$

as $n \to \infty$. ◇

Theorem 2.B.4. *Sørensen (2012, Theorem 1.60) Suppose that $G_n(\theta)$ satisfies the conditions of Theorem 2.B.2, and let δ_n be a sequence of non-zero numbers with $\delta_n \to \infty$ as $n \to \infty$. Suppose that there exists a real-valued, non-degenerate random variable $G(\theta_0)$, such that*

$$\begin{pmatrix} \delta_n G_n(\theta_0) \\ \partial_\theta G_n(\theta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} G(\theta_0) \\ B(\theta_0;\theta_0) \end{pmatrix} .$$

Then, for any consistent G_n-estimator $\hat{\theta}_n$,

$$\delta_n(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} -B(\theta_0;\theta_0)^{-1} G(\theta_0) .$$

◇

Lemma 2.B.5. *Suppose that the conditions of Theorem 2.B.2 hold. Then, for a consistent G_n-estimator $\hat{\theta}_n$,*

$$\hat{\theta}_n = \theta_0 - B(\theta_0,\theta_0)^{-1} G_n(\theta_0) + |\hat{\theta}_n - \theta_0| \varepsilon_n(\theta_0) ,$$

where $|\varepsilon_n(\theta_0)| \xrightarrow{\mathcal{P}} 0$. ◇

CHAPTER 3

Efficient Estimation for Diffusions With Jumps Sampled at High Frequency Over an Increasing Time Interval

---

Nina Munkholt Jakobsen & Michael Sørensen

Department of Mathematical Sciences
University of Copenhagen

Abstract

This paper concerns parametric estimation for ergodic univariate diffusion processes with finite activity jumps, given by stochastic differential equations. The processes are assumed to be observed at high frequency over an increasing time interval, with terminal sampling time going to infinity. It is established that under quite general assumptions, approximate martingale estimating functions yield consistent estimators of the parameters of the process. These estimators are asymptotically normally distributed, and their asymptotic variances may be estimated consistently. In particular, the estimators are rate optimal for drift-jump related parameters. Conditions for rate optimality of estimators of the diffusion parameter, and efficiency of estimators of the drift-jump and diffusion parameters are given in three special cases. The overall conclusion is that, depending on the jump dynamics of the model, it can be considerably more difficult to achieve rate optimal estimators of the diffusion parameter for a jump-diffusion, than for the corresponding continuous diffusion. For rate optimal estimators of the diffusion parameter, the supplementary condition for efficiency is identical to the one for continuous diffusions. Efficiency of estimators of drift-jump parameters essentially requires the following. The relevant coordinate functions of the estimating function must be able to discriminate asymptotically between observations of the process at jump times and non-jump times respectively. In the former case, the coordinate functions are determined by the score function corresponding to the jump. In the latter case, the coordinate functions must behave like an efficient estimating function for the drift parameter of the corresponding continuous diffusion.

3.1 Introduction

In many fields, when modelling phenomena in continuous time, diffusions with jumps are seen as a natural extension or improvement to continuous diffusion processes with Wiener noise or to pure-jump processes. See, e.g. Giraudo and Sacerdote (1997), Jahn et al. (2011), Musila and Lánský (1991), and Patel and Kosko (2008) for some examples from neuroscience, and Kou (2002), De Jong et al. (2001), and Merton (1976) for some applications in finance.

Statistical inference for diffusions with jumps contains a broad spectrum of intriguing challenges. The models have continuous-time dynamics but, while continuous-time sampling is ideal in theory, it is generally not feasible. As is the case for continuous diffusions, a closed-form expression for the likelihood function based on discrete-time observations is usually not available, rendering maximum likelihood estimation somewhat impracticable. However, the presence of jumps also creates a new, crucial obstacle for alternate estimation procedures. To the extent that knowledge of jump times and sizes is needed, it has to be inferred from the discrete-time observations whether one or more jumps are likely to have occurred between any two consecutive observation times, and, if so, how much of the observed increment is attributable to the jump(s).

A multitude of estimation approaches exist in the literature, a non-exhaustive list of references includes the following. In the context of parametric estimation, pseudo-likelihood methods involving primarily Gaussian-inspired approximations of the log-likelihood (or score) function were considered by e.g. Masuda (2011, 2013), Ogihara and Yoshida (2011), Shimizu (2006b), and Shimizu and Yoshida (2006). Closed-form expansion of the transition densities was investigated by e.g. Filipović et al. (2013) and Yu (2007), while Mai (2014) approximated the maximum likelihood estimators obtained from the continuous-time likelihood function. Mancini (2004) proposed a quadratic variation-inspired estimation method in a semiparametric setting, while simulation-based methods were considered by e.g. Giesecke and Schwenkler (2014), and Stramer et al. (2010). Finally, a selection of non-parametric procedures based on discrete observations exist as well, see e.g. Bandi and Nguyen (2003), Mancini (2009), Mancini and Renò (2011), Schmisser (2014) and Shimizu (2006a, 2008, 2009).

This paper concerns parametric estimation in a framework where the ergodic stochastic process $\mathbf{X} = (X_t)_{t\geq 0}$ is a càdlàg solution to a stochastic differential equation of the form

$$dX_t = a(X_t;\theta)\,dt + b(X_t;\theta)\,dW_t + \int_{\mathbb{R}} c(X_{t-},z;\theta)\,N^\theta(dt,dz)\,. \tag{3.1.1}$$

The drift and diffusion coefficients a and b, and the jump coefficient c are known, deterministic functions of $(y;\theta)$ and $(y,z;\theta)$ respectively, and θ is the unknown, finite-dimensional parameter to be estimated. As usual, $\mathbf{X}_- = (X_{t-})_{t\geq 0}$ is defined as the process of left limits of $\mathbf{X}$. The standard Wiener process $(W_t)_{t\geq 0}$ is supposed to be independent of $N^\theta(dt,dz)$, a time-homogeneous Poisson random measure on $[0,\infty)\times\mathbb{R}$, with the intensity measure μ_θ given by $\mu_\theta(dt,dz) = \nu_\theta(dz)\,dt$. Furthermore, ν_θ is a Lévy measure on $\mathbb{R}$ for which

$\nu_\theta(\mathbb{R}) < \infty$, i.e. the jumps of $\mathbf{X}$ are of finite activity. For simplicity, X_t is assumed to be one-dimensional.

Let $(\Delta_n)_{n\in\mathbb{N}}$ be a sequence of strictly positive numbers. For each $n \in \mathbb{N}$, we assume observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$ of $\mathbf{X}$ over the interval $[0, n\Delta_n]$, at discrete, equidistant time-points $t_i^n = i\Delta_n$, for $i = 0, 1, \ldots, n$. Asymptotics are considered as $n \to \infty$, in which case it is assumed that $\Delta_n \to 0$ and $n\Delta_n \to \infty$. With this observation scheme, $\mathbf{X}$ is said to be observed at high frequency, over an increasing time interval, with terminal sampling time $t_n^n = n\Delta_n$ going to infinity. In the limit, the whole sample path of the process is (hypothetically) observed, with full information about jump times and sizes.

Local asymptotic normality (LAN), and, for fixed-interval asymptotics, local asymptotic mixed normality (LAMN) results are an active area of research for stochastic processes with jumps, recent developments including Becheri et al. (2014), Clément and Gloter (2015), Kawai and Masuda (2013) and Kohatsu-Higa et al. (2014, 2015). Within the context of local asymptotic normality, it is quite straightforward to characterise rate optimality and efficiency of estimators. In the absence of general local asymptotic normality results for the present setup, the criteria for rate optimality and efficiency used here are more heuristic in nature, motivated not only by the applicable local asymptotic normality results of Becheri et al., and Kohatsu-Higa et al., but also by results of Gobet (2002), Shimizu and Yoshida (2006), and Sørensen (1991).

Parametric estimation situations similar that described above were considered by e.g. Ogihara and Yoshida (2011), and Shimizu and Yoshida (2006), in the case of finite-activity jumps, and Masuda (2011, 2013), and Shimizu (2006b), who also allowed infinite-activity jumps.[1] Shimizu and Yoshida proposed a technique to judge whether or not a jump is likely to have occurred between two observation times t_{i-1}^n and t_i^n. They used this technique to create a contrast function for estimation in the sub-model of (3.1.1) given by

$$dX_t = a(X_t; \alpha)\, dt + b(X_t; \beta)\, dW_t + \int_{\mathbb{R}} c(X_{t-}, z; \alpha)\, N^{\alpha}(dt, dz)\,, \tag{3.1.2}$$

where the general parameter θ is split into a drift-jump parameter α and a separate diffusion parameter β. Their contrast function treats the pair $(X_{t_{i-1}^n}, X_{t_i^n})$ differently, depending on whether or not a jump is presumed to have occurred between the two observation times. Shimizu and Yoshida argued that estimators based on their contrast function are (rate optimal and) efficient for the drift-jump parameter, and, by the criteria laid down in the present paper, the same goes for the diffusion parameter. The contrast function used by Ogihara and Yoshida (2011) was almost identical to that of Shimizu and Yoshida, while the estimating function used by Shimizu (2006b) was heavily inspired it.

Masuda considered estimation within a class of stochastic differential equation models with jumps, which, in special cases, overlap with sub-models of (3.1.1) of the form

$$dX_t = a(X_t; \alpha)\, dt + b(X_t; \beta)\, dW_t + \int_{\mathbb{R}} \tilde{c}(X_{t-}, \beta) z\, N(dt, dz)\,.$$

[1] Several of these papers assumed multivariate processes as well, in the following, we only refer to their results in the univariate case.

In the articles of Masuda (2011, 2013), estimation was performed using a specific type of Gaussian quasi-likelihood functions. As noted by Masuda, a type of estimation known to work well for diffusions without jumps. Among other things, Masuda studied the asymptotic properties of his Gaussian quasi-likelihood estimators under the current asymptotic scenario. He pointed out that in the presence of jumps, these estimators are not efficient for the drift or diffusion-jump parameters of the model, or even rate optimal for parameters of the diffusion coefficient.

Under certain regularity conditions, the Gaussian quasi-likelihood estimators investigated by Masuda fit into the framework of approximate martingale estimating functions, the topic of this paper. Approximate martingale estimating functions, which can be viewed as approximations to the score function, may be written on the form

$$G_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \,. \tag{3.1.3}$$

For some constant $\kappa \geq 2$, the $\mathbb{R}^d$-valued function $g(t, y, x; \theta)$ satisfies a conditional expectation condition of the form

$$\mathbb{E}_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}) = \Delta_n^\kappa R_\theta(\Delta_n, X_{t_{i-1}^n}) \,, \tag{3.1.4}$$

with a remainder term on the right-hand side which can be controlled as necessary. Estimators are essentially obtained as solutions to the estimating equation $G_n(\theta) = 0$. More precise definitions of approximate martingale estimating functions and the corresponding estimators are given in Section 3.2.3. Estimating functions of this type were also used by, e.g. Bibby and Sørensen (1995), Jacobsen (2001, 2002), Sørensen (2010) and Uchida (2004) for continuous diffusions.[2] To our knowledge, high-frequency asymptotics for the general class of approximate martingale estimating functions have not previously been studied for diffusions with jumps.

The observation scheme considered here matches that of Sørensen (2010). For continuous diffusions of the form (3.1.2) with $c(x, z; \alpha) \equiv 0$, Sørensen showed that under simple conditions, approximate martingale estimating function-based estimators of the drift and diffusion parameters α and β are rate optimal and efficient. Sørensen also argued that the theory of approximate martingale estimating functions covers a number of other estimators proposed in the literature on continuous diffusions. On the one hand, estimators which are efficient under the present asymptotic scenario, e.g. those of Florens-Zmirou (1989), Kessler (1997) and Yoshida (1992), and, on the other hand, a number of estimators which perform well under other sampling schemes (see Sørensen (2010) for further references).

Based on their efficacy in the case of continuous diffusions, we believe that a thorough investigation into the behaviour of general approximate martingale estimating functions in the current setting is justified, and has the potential to contribute valuable information

[2]Approximate martingale estimating functions were also applied in the setting of continuous diffusions in Chapter 2.

to the field of estimation for diffusions with jumps. For example, rate optimality and efficiency conditions in the style of Sørensen (2010), but for diffusions with jumps, could perhaps facilitate the construction of efficient approximate martingale estimating functions in sub-models of (3.1.1). The results presented in this paper may be considered preliminary findings on the matter, with much research yet to be done.

Initially, supposing the existence of a true parameter θ_0, we provide the general Theorem 3.3.2, which establishes existence and uniqueness properties, and asymptotic distributions for consistent estimators of θ_0 based on approximate martingale estimating functions. In general terms, the theorem states that under suitable regularity assumptions on the jump-diffusion model (3.1.1) and on the chosen approximate martingale estimating function (3.1.3), there exists a consistent estimator $\hat{\theta}_n$, such that

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, V(\theta_0)) . \tag{3.1.5}$$

$\mathcal{N}_d(0, V)$ denotes the d-dimensional zero-mean Gaussian distribution with variance V, and $\xrightarrow{\mathcal{D}}$ denotes convergence in distribution under the true probability measure. Furthermore, $V(\theta_0)$ may be estimated consistently. For (exact) martingale estimating functions, in which case the right-hand side of (3.1.4) vanishes, there are no additional requirements on the speed at which Δ_n goes to zero. For all other approximate martingale estimating functions it is required that $n\Delta_n^{2\kappa-1} \to 0$, with κ determined by (3.1.4).

Adapting the model and estimating function considered by Masuda (2011) to our framework and assumptions, it is seen in Example 3.3.3 that the limit distribution in (3.1.5) relates to the one obtained by Masuda (2011, Theorem 3.4). Similarly, the limit distribution is comparable to the one obtained by Masuda (2013, Theorem 2.9).

Having established the general theorem, we pursue the question of rate optimality and efficiency within the sub-model (3.1.2). As approximate martingale estimating functions are not a priori designed to discriminate observed increments with jumps from those without, we expect such a distinguishing mechanism to be an inherent feature of the conditions for rate optimality and efficiency, to the extent that it is necessary.

Preceded by some extra regularity assumptions not mentioned here, we state conditions under which an approximate martingale estimating function yields rate optimal and efficient estimators in three sub-models of (3.1.2). The first model is assumed to have only an unknown, d-dimensional drift-jump parameter α, the second only an unknown, one-dimensional diffusion parameter β, and the third a two-dimensional drift-jump parameter α and a one-dimensional diffusion parameter β, both unknown. In order to obtain rate optimality of the estimator of β when using non-exact martingale estimating functions, it is assumed that $n\Delta_n^{2(\kappa-1)} \to 0$.

In addition to the rate optimality and efficiency conditions obtained by Sørensen (2010) for diffusions without jumps, several new jump-related conditions appear. In particular, an important observation is made in connection with the conditions for efficient estimation of the drift-jump parameter α. In the limit $\Delta_n \to 0$, when a full sample path of $\mathbf{X}$ is (hypothetically) observed, and all jump times and sizes may be identified by $X_t \neq X_{t-}$, the following is required: In general terms, whenever $g(0, X_t, X_{t-}; \theta)$ is evaluated at a jump time t, particular coordinate functions should behave like the score function of the distribution of the jump. At all other times, these coordinate functions should behave like those of an efficient estimating function for the drift parameter of the corresponding continuous diffusion. In other words, not only should these coordinates of the estimating function be able to discriminate, asymptotically, between pairs $(y, x) = (X_t, X_{t-})$ with $X_t \neq X_{t-}$ and $X_t = X_{t-}$, there is essentially no freedom of choice regarding the coordinate functions in the former case.

In connection with our rate optimality conditions for estimators of the diffusion parameter, the following is observed as well. For models with certain types of finite activity jump dynamics, creating an estimating function which is rate optimal for the diffusion parameter, and which satisfies the remaining regularity assumptions we impose on the function, might be quite challenging, and sometimes impossible. This stands in contrast to the situation for continuous diffusions studied by Sørensen (2010)[3], where it is quite straightforward to construct rate optimal estimating functions satisfying essentially the same regularity assumptions as here.

Finally, as a suggestion for further research, we discuss how, in certain models, an approximate martingale estimating function satisfying the rate optimality and efficiency conditions put forth might be constructed as a modification of the efficient contrast function of Shimizu and Yoshida (2006).

The general method of proof in this paper is inspired by that of Sørensen (2010). However, the presence of jumps complicates matters considerably, and creates a large variety of additional, complex challenges to deal with.

The structure of the rest of this paper is as follows: Section 3.2 presents definitions, notation and terminology used throughout the paper, as well as the main assumptions imposed on the jump-diffusion and the approximate martingale estimating functions. Section 3.3 presents the general theorem on approximate martingale estimating function-based estimators of the parameter of the jump-diffusion model (3.1.1). Section 3.4 is devoted to the question of rate optimality and efficiency of estimators of the drift-jump and diffusion parameters in sub-models of the form (3.1.2). In particular, our criteria for rate optimality and efficiency are elaborated on. Section 3.5 contains main lemmas used to prove our theorems, the proofs of these theorems, and the proofs of the main lemmas. Appendix 3.A consists of a considerable number of technical auxiliary results used in the proofs of these main lemmas, most of them presented with a proof. Appendix 3.B summarises some important theorems from the literature without proofs.

[3] As well as the situation for continuous diffusions studied in Chapter 2 of this thesis.

3.2 Preliminaries

Section 3.2.1 elaborates on, and serves to introduce some notation associated with the jump-diffusion process and the observation scheme under consideration. In Section 3.2.2, a notation and terminology regarding the concept of polynomial growth is established for subsequent use. Section 3.2.3 contains formal definitions of approximate martingale estimating functions and their corresponding estimators. Section 3.2.4 introduces the general assumptions on the jump-diffusion processes (Assumption 3.2.5) and on the estimating functions (Assumption 3.2.6). Finally, in Section 3.2.5, notation pertaining to the (infinitesimal) generator of the diffusion process is established, and some useful technical results expressed in terms of the generator are discussed.

For a moment, let $p, q \in \mathbb{N}$. In this paper, the following notation is used: $M^\star$ denotes transposition of a matrix (or vector) M, and $\|M\|$ the Euclidean norm. For any $\mathbb{R}^p$-valued function f, let $f = (f_1, \ldots, f_p)^\star$, where f_j denotes the j'th (real-valued) coordinate function of f. For an $\mathbb{R}^q$-valued argument u, let $\partial_{u_k} f_j$ be the jk'th element of the $p \times q$ matrix $\partial_u f$, where $\partial_{u_k} f_j$ denotes the (partial) derivative of f_j with respect to u_k. Furthermore, let $f^2 = (f_1^2, \ldots, f_d^2)^\star$. For a $p \times q$ matrix-valued function $F = (F_{jk})$ (with real-valued coordinate functions), we define $\partial_u F = (\partial_u F_{jk})$, if u is real-valued, and $F^2 = (F_{ij}^2)$.

3.2.1 Model and Observations

Let $(\Omega, \mathcal{F})$ be a measurable space equipped with a filtration $(\mathcal{F}_t)_{t\geq 0}$ and a family of probability measures $(\mathbb{P}_\theta)_{\theta\in\Theta}$. The d-dimensional parameter set Θ is assumed to contain the true parameter θ_0. Assume also an $(\mathcal{F}_t)_{t\geq 0}$-adapted standard Wiener process $\mathbf{W} = (W_t)_{t\geq 0}$, and an independent, time-homogeneous Poisson random measure $N^\theta(dt, dz)$ on $[0, \infty)\times\mathbb{R}$, with the intensity measure μ_θ given by $\mu_\theta(dt, dz) = \nu_\theta(dz)\, dt$. For all $\theta \in \Theta$, ν_θ is a Lévy measure on $\mathbb{R}$, which satisfies that $\nu(\{0\}) = 0$ and $\nu_\theta(\mathbb{R}) < \infty$.

Consider the stochastic differential equation

$$dX_t = a(X_t; \theta)\, dt + b(X_t; \theta)\, dW_t + \int_{\mathbb{R}} c(X_{t-}, z; \theta)\, N^\theta(dt, dz)\,, \quad X_0 = U\,, \tag{3.2.1}$$

where U is an $\mathcal{F}_0$-measurable random variable, and independent of $\mathbf{W}$ and N^θ. It is assumed that X_t takes its values in an open (not necessarily bounded) interval $\mathcal{X} \subseteq \mathbb{R}$, and that the drift, diffusion and jump coefficients, $a, b : \mathcal{X}\times\Theta \to \mathbb{R}$ and $c : \mathcal{X}\times\mathbb{R}\times\Theta \to \mathbb{R}$ respectively are known, deterministic functions.

The assumption $\nu_\theta(\mathbb{R}) < \infty$ implies that the jumps of $\mathbf{X}$ have *finite activity*, i.e. that there are ($\mathbb{P}_\theta$-almost surely) only finitely many jumps in any given finite time interval $I \subseteq [0, \infty)$. Consequently, the stochastic integral in (3.2.1) is well-defined. Under $\mathbb{P}_\theta$, X_t may be written as

$$X_t = X_0 + \int_0^t a(X_s; \theta)\, ds + \int_0^t b(X_s; \theta)\, dW_s + \int_0^t \int_{\mathbb{R}} c(X_{s-}, z; \theta)\, N^\theta(ds, dz)\,.$$

From (3.2.1) it is seen that in intervals with no jumps, $\mathbf{X}$ follows the dynamics of the corresponding continuous diffusion with $c(x,z;\theta) \equiv 0$.

Let $(\Delta_n)_{n\in\mathbb{N}}$ be a sequence of strictly positive numbers such that

$$\Delta_n \to 0 \quad \text{and} \quad n\Delta_n \to \infty \quad \text{as} \quad n \to \infty\,,$$

with $\Delta_0 = \max\{\Delta_n : n \in \mathbb{N}\}$. For each $n \in \mathbb{N}$, $\mathbf{X}$ is supposed to be sampled equidistantly over the time-interval $[0, n\Delta_n]$ at times $t_i^n = i\Delta_n$, $i = 0, 1, \ldots, n$, yielding the observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$. Define $\mathcal{G}_{n,i}$, $i = 1, 2, \ldots, n$, to be the σ-algebra generated by $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_i^n})$, and let $\mathcal{G}_n = \mathcal{G}_{n,n}$.

3.2.2 Polynomial Growth

Throughout this paper, in order to avoid cumbersome notation, C denotes a generic, strictly positive, real-valued constant. Often, the notation C_u is used to emphasise that the constant depends on some u, where u may be, e.g. a parameter-value $\theta \in \Theta$, some number $m \in \mathbb{N}_0$, a set $K \subseteq \Theta$ or a combination of these. It is important to note that, for example, in an expression of the form $C_u(1 + |x|^{C_u})$, the factor C_u and the exponent C_u need not be equal. C or C_u often depend (implicitly) on, e.g. the unknown parameter θ_0, the maximum time step Δ_0 and the dimension d of the parameter space Θ, but never on the sample size n.

Definition 3.2.1 (Polynomial Growth)**.** A (coordinate) function $f : \mathcal{X}^2 \times \Theta \to \mathbb{R}$ is of polynomial growth in x and y, if for each $\theta \in \Theta$ there exist constants $C_\theta > 0$ such that

$$|f(y, x; \theta)| \leq C_\theta(1 + |x|^{C_\theta} + |y|^{C_\theta})$$

for $x, y \in \mathcal{X}$.

Choose $\varepsilon_0 > 0$ and define $(0, \Delta_0)_{\varepsilon_0} = (0 - \varepsilon_0, \Delta_0 + \varepsilon_0)$. Then, a function $f : (0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta \to \mathbb{R}$ is of polynomial growth in x and y, uniformly for $t \in (0, \Delta_0)_{\varepsilon_0}$ and θ in compact, convex sets, if for each compact, convex set $K \subseteq \Theta$, there exist constants $C_K > 0$ such that

$$\sup_{t\in(0,\Delta_0)_{\varepsilon_0},\, \theta\in K} |f(t, y, x, \theta)| \leq C_K \left(1 + |x|^{C_K} + |y|^{C_K}\right)$$

for $x, y \in \mathcal{X}$.

$C^{\text{pol}}_{p,q,r,s}((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ denotes the class of real-valued functions $f(t, y, x, \theta)$ which satisfy that

(i) f and the mixed partial derivatives $\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l f(t, y, x; \theta)$, $i = 0, \ldots, p$, $j = 0, \ldots, q$, $k = 0, \ldots, r$, $l = 0, \ldots, s$ and $m = 1, \ldots, d$, exist and are continuous on $(0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta$.

(ii) f and all the mixed partial derivatives from (i) are of polynomial growth in x and y, uniformly for $t \in (0, \Delta_0)_{\varepsilon_0}$ and θ in compact, convex sets.

Similarly, the classes $C^{\text{pol}}_{p,r,s}((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X} \times \Theta)$, $C^{\text{pol}}_{q,r,s}(\mathcal{X}^2 \times \Theta)$, $C^{\text{pol}}_{q,s}(\mathcal{X} \times \Theta)$ and $C^{\text{pol}}_{q}(\mathcal{X})$ are defined for functions of the form $f(t, x; \theta)$, $f(y, x; \theta)$, $f(y; \theta)$ (or $f(x; \theta)$) and $f(y)$ (or $f(x)$). ◇

Definition 3.2.2 (Product-Polynomial Growth)**.** For functions of the form $f : \mathcal{X} \times \mathbb{R} \times \Theta \to \mathbb{R}$, $f(y, z; \theta)$ is defined to be of product-polynomial growth in y and z, uniformly for θ in compact, convex sets if for each of such sets $K \subseteq \Theta$, there exist constants $C_K > 0$ so that

$$\sup_{\theta \in K} |f(y, z, \theta)| \leq C_K \left(1 + |y|^{C_K}\right)\left(1 + |z|^{C_K}\right)$$

for all $y \in \mathcal{X}$ and $z \in \mathbb{R}$.

$C^{\text{p-pol}}_{q,s}(\mathcal{X} \times \mathbb{R} \times \Theta)$ denotes the class of real-valued functions $f(y, z; \theta)$ which satisfy that

(i) f and the mixed partial derivatives $\partial^i_y \partial^j_{\theta_k} f(y, z; \theta)$, $i = 0, \ldots, q$, $j = 0, \ldots, s$, and $k = 1, \ldots, d$, exist and are continuous on $\mathcal{X} \times \mathbb{R} \times \Theta$.

(ii) f and all the mixed partial derivatives from (i) are of product-polynomial growth in y and z, uniformly for θ in compact, convex sets. ◇

Note that in Definition 3.2.2, differentiability of f with respect to z is not required. For functions not depending on t (respectively, θ), the "uniformly for t" ("uniformly for θ") parts of Definitions 3.2.1 and 3.2.2 become superfluous.

For the duration of this paper, $R(t, y, x; \theta)$ denotes a generic function defined on the set $(0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta$, which may be real-valued, $\mathbb{R}^d$-valued, or take values in the space of $d \times d$ matrices with real entries. The coordinate functions of $R(t, y, x; \theta)$ are of polynomial growth in x and y, uniformly for $t \in (0, \Delta_0)_{\varepsilon_0}$ and θ in compact, convex sets. $R(t, y, x, \theta)$ may depend (implicitly) on θ_0. $R(t, x; \theta)$, $R(y, x; \theta)$ and $R(t, x)$ are defined correspondingly. Finally, e.g. $R_\lambda(t, x; \theta)$ indicates that $R(t, x; \theta)$ also depends on $\lambda \in \Theta$ in an unspecified way. In particular, $R_\theta(t, x, \theta) = R_\theta(t, x)$.

3.2.3 Approximate Martingale Estimating Functions

Let $\mathbb{E}_\theta$ denote expectation under $\mathbb{P}_\theta$. In this paper, (approximate) martingale estimating functions, along the lines of those defined by, e.g. Sørensen (2012, Sections 1.3 & 1.5.3), are defined as follows:

Definition 3.2.3. Let $g(t, y, x; \theta)$ be an $\mathbb{R}^d$-valued function defined on $(0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta$, with $(0, \Delta_0)_{\varepsilon_0} = (0 - \varepsilon_0, \Delta_0 + \varepsilon_0)$ for some $\varepsilon_0 > 0$. Suppose that there exists some constant $\kappa \geq 2$, such that

$$\mathbb{E}_\theta\left(g(\Delta_n, X_{t^n_i}, X_{t^n_{i-1}}; \theta) \mid X_{t^n_{i-1}}\right) = \Delta^\kappa_n R_\theta(\Delta_n, X_{t^n_{i-1}}) \tag{3.2.2}$$

for all $n \in \mathbb{N}$, $i = 1, \ldots, n$ and $\theta \in \Theta$. Then, the function

$$G_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^{n} g(\Delta_n, X_{t^n_i}, X_{t^n_{i-1}}; \theta) \tag{3.2.3}$$

is referred to as an *approximate martingale estimating function*. In particular, when (3.2.2) is satisfied with $R_\theta(t,x) \equiv 0$, (3.2.3) is referred to as a *martingale estimating function*. $\diamond$

When not ambiguous, approximate martingale estimating functions may sometimes just be referred to as *estimating functions* in the following. By the Markov property of $\mathbf{X}$, it is seen that when $R_\theta(t,x) \equiv 0$, $(G_{n,i})_{1\le i\le n}$ defined by

$$G_{n,i}(\theta) = \frac{1}{n\Delta_n}\sum_{j=1}^{i} g(\Delta_n, X_{t_j^n}, X_{t_{j-1}^n};\theta)$$

is a zero-mean, $\mathbb{R}^d$-valued $(\mathcal{G}_{n,i})_{1\le i\le n}$-martingale under $\mathbb{P}_\theta$ for each $n \in \mathbb{N}$, thus giving rise to the terminology in Definition 3.2.3. An approximate martingale estimating function is essentially an approximation to the score function of the observations $(X_{t_0^n}, X_{t_1^n}, \dots, X_{t_n^n})$, conditional on $X_{t_0^n}$, which itself is a martingale.

A *G_n-estimator* $\hat{\theta}_n$, that is, an estimator based on the approximate martingale estimating function $G_n(\theta)$, is essentially obtained as a solution to the estimating equation $G_n(\theta) = 0$. A more precise definition, based on Jacod and Sørensen (2012, Definition 2.1) and Sørensen (2012, Definition 1.57), is given in Definition 3.2.4.

Formally, an approximate martingale estimating function may be considered a function of both $\theta \in \Theta$ and $\omega \in \Omega$, while a G_n-estimator may be considered a function of ω. For the purpose of the following definition, it is convenient to make this dependence explicit and write $G_n(\theta,\omega)$ and $\hat{\theta}_n(\omega)$.

Definition 3.2.4. Let $G_n(\theta,\omega)$ be an approximate martingale estimating function as defined in Definition 3.2.3. Put $\Theta_\infty = \Theta \cup \{\infty\}$ and let

$$D_n = \{\omega \in \Omega \mid G_n(\theta,\omega) = 0 \text{ has at least one solution } \theta \in \Theta\}.$$

A *G_n-estimator* $\hat{\theta}_n(\omega)$ is any $\mathcal{G}_n$-measurable function $\Omega \to \Theta_\infty$ which satisfies that for $\mathbb{P}_{\theta_0}$-almost all ω, $\hat{\theta}_n(\omega) \in \Theta$ and $G_n(\hat{\theta}_n(\omega),\omega) = 0$ if $\omega \in D_n$, and $\hat{\theta}_n(\omega) = \infty$ if $\omega \notin D_n$. $\diamond$

For any invertible $d \times d$ matrix M_n with real entries, which may depend on e.g. Δ_n, $G_n(\theta)$ and $M_nG_n(\theta)$ yield identical estimators of θ. The estimating functions $G_n(\theta)$ and $M_nG_n(\theta)$ are referred to as *versions* of each other. For any given estimating function, it is sufficient that there exists a version of the function which satisfies the assumptions of this paper, in order to draw conclusions about the resulting estimators.

3.2.4 Assumptions

In the following, $\xrightarrow{\mathcal{P}}$ denotes convergence in probability. Unless otherwise mentioned, it is assumed to be under $\mathbb{P}_{\theta_0}$ as $n \to \infty$.

Assumption 3.2.5. *The parameter set Θ is a non-empty, open, not necessarily bounded subset of $\mathbb{R}^d$ for some $d \in \mathbb{N}$, which contains the true parameter θ_0. The càdlàg, $(\mathcal{F}_t)$-adapted Markov process $\mathbf{X} = (X_t)_{t\ge 0}$ solves a stochastic differential equation of the form*

(3.2.1), the coefficients of which satisfy that

$$a(y;\theta),\ b(y;\theta) \in C^{pol}_{2,2}(\mathcal{X} \times \Theta) \quad \text{and} \quad c(y,z;\theta) \in C^{p\text{-}pol}_{2,2}(\mathcal{X} \times \mathbb{R} \times \Theta)\ .$$

The following holds for all $\theta \in \Theta$:

(i) *For all $y \in \mathcal{X}$, $b^2(y;\theta) > 0$.*

(ii) *There exist real-valued constants $C_\theta > 0$ such that for all $x, y \in \mathcal{X}$ and $z \in \mathbb{R}$,*

$$|a(x;\theta) - a(y;\theta)| + |b(x;\theta) - b(y;\theta)| + |c(x,z;\theta) - c(y,z;\theta)|(1 + |z|^{C_\theta})^{-1} \leq C_\theta |x - y|\ .$$

(iii) *There exist real-valued constants $C_\theta > 0$ such that*

$$|a(y;\theta)| + |b(y;\theta)| + |c(y,z;\theta)|(1 + |z|^{C_\theta})^{-1} \leq C_\theta(1 + |y|)$$

for all $x, y \in \mathcal{X}$ and $z \in \mathbb{R}$.

(iv) *For all $m \in \mathbb{N}$,*

$$\sup_{t \in [0,\infty)} \mathbb{E}_\theta\left(|X_t|^m\right) < \infty\ .$$

(v) **X** *is ergodic, i.e. there exists an invariant probability measure π_θ such that for any π_θ-integrable function f,*

$$\frac{1}{T}\int_0^T f(X_t)\,dt \xrightarrow{\mathcal{P}} \int_{\mathcal{X}} f(x)\,\pi_\theta(dx) \tag{3.2.4}$$

under $\mathbb{P}_\theta$ as $T \to \infty$. Also, for all $m \in \mathbb{N}$,

$$\int_{\mathcal{X}} |x|^m\,\pi_\theta(dx) < \infty\ .$$

(vi) *The Lévy measure ν_θ has density $q(z;\theta) = \xi(\theta)p(z;\theta)$ with respect to a σ-finite measure $\tilde{\nu}$, where $p(z;\theta)$ is a probability density with respect to $\tilde{\nu}$.*

Finally, the following holds for the densities of the Lévy measures:

(vii) *The functions $\theta \mapsto \partial^j_{\theta_k} q(z;\theta)$, $j = 0, 1, 2$, $k = 1, \ldots, d$, exist and are continuous, and for each compact, convex set $K \subseteq \Theta$, there exists a measurable function $\varphi_K : \mathbb{R} \to [0,\infty)$ with*

$$\int_{\mathbb{R}} |z|^m\,\varphi_K(z)\,\tilde{\nu}(dz) < \infty$$

for all $m \in \mathbb{N}_0$, such that for all $z \in \mathbb{R}$ and $\theta \in K$,

$$q(z;\theta) + \sum_{k=1}^d |\partial_{\theta_k} q(z;\theta)| + \sum_{k=1}^d |\partial^2_{\theta_k} q(z;\theta)| \leq \varphi_K(z)\ .$$

◇

Note that by Assumption 3.2.5.(vii), for all $\theta \in \Theta$ and $m \in \mathbb{N}$,

$$\int_{\mathbb{R}} |z|^m \, \nu_\theta(dz) < \infty \,. \tag{3.2.5}$$

Assumption 3.2.5.(ii) implies that $a(y;\theta)$, $b(y;\theta)$ and $c(y,z;\theta)$ are Lipschitz continuous in y. Under Assumption 3.2.5, there exist constants $C_\theta > 0$ such that

$$\int_{\mathbb{R}} |c(x,z;\theta) - c(y,z;\theta)|^2 \, \nu_\theta(dz) \le C_\theta \int_{\mathbb{R}} \left(1 + |z|^{C_\theta}\right) \nu_\theta(dz) \, |x-y|^2 \le C_\theta \, |x-y|^2$$

$$\int_{\mathbb{R}} c^2(y,z;\theta) \, \nu_\theta(dz) \le C_\theta \int_{\mathbb{R}} \left(1 + |z|^{C_\theta}\right) \nu_\theta(dz) \, (1+|y|)^2 \le C_\theta(1+|y|^2)$$

for all $x, y \in \mathcal{X}$, from which it follows that conditions C1 and C2 of Applebaum (2009, pp. 365-366) are satisfied. Thus, by Applebaum (2009, Theorems 6.2.9 & 6.4.6), there exists a unique, càdlàg, $(\mathcal{F}_t)$-adapted (strong) solution to (3.2.1) under each $\mathbb{P}_\theta$, which is also a Markov process. That is, $\mathbf{X}$ is well-defined.

By Assumption 3.2.5.(iii), for all $\theta \in \Theta$ there exist constants $C_\theta > 0$ such that

$$\int_{\mathbb{R}} |c(y,z;\theta)| \, \nu_\theta(dz) \le C_\theta(1+|y|) \int_{\mathbb{R}} (1 + |z|^{C_\theta}) \, \nu_\theta(dz) \le C_\theta(1+|y|) \,,$$

which means that $\tilde{a}(y;\theta)$ given by

$$\tilde{a}(y;\theta) = a(y;\theta) + \int_{\mathbb{R}} c(y,z;\theta) \, \nu_\theta(dz)$$

is also of linear growth in y. Sometimes, under $\mathbb{P}_\theta$, it is convenient to write (3.2.1) as

$$dX_t = \tilde{a}(X_t;\theta) \, dt + b(X_t;\theta) \, dW_t + \int_{\mathbb{R}} c(X_{t-},z;\theta) \, (N^\theta - \mu_\theta)(dt,dz) \,, \quad X_0 = U \tag{3.2.6}$$

and X_t as

$$X_t = X_0 + \int_0^t \tilde{a}(X_s;\theta) \, ds + \int_0^t b(X_s;\theta) \, dW_s + \int_0^t \int_{\mathbb{R}} c(X_{s-},z;\theta) \, (N^\theta - \mu_\theta)(ds,dz) \,. \tag{3.2.7}$$

Assumption 3.2.5 is similar to assumptions of e.g. Masuda (2013), Ogihara and Yoshida (2011), and Shimizu and Yoshida (2006). E.g. Masuda (2007, 2008) gives conditions that ensure the existence of an ergodic theorem of the form (3.2.4), and under which $\mathbf{X}$ has bounded moments as in Assumption 3.2.5.(iv).

Assumption 3.2.6. *For some interval* $(0,\Delta_0)_{\varepsilon_0} = (0-\varepsilon_0, \Delta_0+\varepsilon_0)$ *with* $\varepsilon_0 > 0$*, the* $\mathbb{R}^d$*-valued function* $g(t,y,x;\theta)$ *satisfies that for* $j = 1,\ldots,d$,

$$g_j(t,y,x;\theta) \in C^{pol}_{1,4,1,2}\left((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta\right) \,,$$

and defines an approximate martingale estimating function $G_n(\theta)$ *as prescribed by Definition 3.2.3. In particular,*

(i) *for some constant $\kappa \geq 2$, and for all $n \in \mathbb{N}$, $i = 1, \dots, n$ and $\theta \in \Theta$,*

$$\mathbb{E}_\theta\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^\kappa R_\theta(\Delta_n, X_{t_{i-1}^n}).$$

Also, the following is true for all $\theta \in \Theta$:

(ii) *The expansion*

$$g(\Delta, y, x; \theta) = g(0, y, x; \theta) + \Delta g^{(1)}(y, x; \theta) + \Delta^2 R(\Delta, y, x; \theta)$$

holds for $\Delta \in (0, \Delta_0)_{\varepsilon_0}$ and $x, y \in \mathcal{X}$, where $g^{(1)} = (g_1^{(1)}, \dots, g_d^{(1)})^\star$, and $g_j^{(1)}(y, x; \theta)$ denotes the 1st partial derivative of $g_j(t, y, x; \theta)$ with respect to t, evaluated in $t = 0$.

◇

The assumptions of polynomial growth, together with the assumptions on the moments of e.g. X_t, ν_θ and π_θ, serve to simplify the exposition and proofs in this paper, and could be relaxed. Likewise, asides from ensuring (3.2.5), the purpose of Assumption 3.2.5.(vii) is to provide sufficient (but not necessary) conditions for interchanging integration and differentiation in Lemma 3.A.2.

3.2.5 The Infinitesimal Generator

Definition 3.2.7. Suppose that Assumption 3.2.5 holds. Let

$$f(t, y, x; \theta) \in C^{\text{pol}}_{0,2,0,0}((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$$

and define, for $\lambda \in \Theta$, the *(infinitesimal) generator* $\mathcal{L}_\lambda$ (through its action on f) by

$$\begin{aligned} &\mathcal{L}_\lambda f(t, y, x; \theta) \\ &\quad = a(y; \lambda)\partial_y f(t, y, x; \theta) + \tfrac{1}{2} b^2(y; \lambda)\partial_y^2 f(t, y, x; \theta) \\ &\qquad + \int_{\mathbb{R}} (f(t, y + c(y, z; \lambda), x; \theta) - f(t, y, x; \theta))\, \nu_\lambda(dz). \end{aligned} \tag{3.2.8}$$

◇

Often, the notation $\mathcal{L}_\lambda f(t, y, x; \theta) = \mathcal{L}_\lambda(f(t; \theta))(y, x)$ is used. Since $\nu_\lambda(\mathbb{R}) < \infty$, Lemma 3.A.1 yields constants $C_{\lambda,\theta} > 0$ such that

$$\int_{\mathbb{R}} |f(t, y + c(y, z; \lambda), x; \theta) - f(t, y, x; \theta))|\, \nu_\lambda(dz) \leq C_{\lambda,\theta}\left(1 + |x|^{C_{\lambda,\theta}} + |y|^{C_{\lambda,\theta}}\right)$$

for $t \in (0, \Delta_0)_{\varepsilon_0}$, $x, y \in \mathcal{X}$ and $\theta \in \Theta$, implying that the integral in (3.2.8) is well-defined. In essence, Lemmas 3.A.1 and 3.A.2 of Appendix 3.A.1 verify that integrals with respect to the Lévy measure inherit polynomial growth properties of the integrand. This is often used (implicitly) in the current paper, in particular in connection with applications of the infinitesimal generator.

The operator $\mathcal{L}_\lambda$, always acting on the variable y of the function it is applied to, is defined correspondingly for e.g. functions $f(y) \in C_2^{\text{pol}}(\mathcal{X})$ and $f(y, x; \theta) \in C_{2,0,0}^{\text{pol}}(\mathcal{X}^2 \times \Theta)$, and functions $f(t, y, x, \mathbf{z}_k; \theta)$, for which

$$((t, y, x; \theta) \mapsto f(t, y, x, \mathbf{z}_k; \theta)) \in C_{0,2,0,0}^{\text{pol}}((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$$

for $\tilde{\nu}$-almost all $\mathbf{z}_k = (z_1, \ldots, z_k)^\star \in \mathbb{R}^k$. In the latter case, the notation $\mathcal{L}_\lambda f(t, y, x, \mathbf{z}_k; \theta) = \mathcal{L}_\lambda(f(t, \mathbf{z}_k; \theta))(y, x)$ is used.

Whenever the expression is well-defined, $\mathcal{L}_\lambda^2 f$ is to be understood as $\mathcal{L}_\lambda(\mathcal{L}_\lambda f)$, and similarly $\mathcal{L}_\lambda^k f = \mathcal{L}_\lambda(\mathcal{L}_\lambda^{k-1} f)$ for $k \in \mathbb{N}$ with $\mathcal{L}_\lambda^0 f = f$. If $f = (f_1, \ldots, f_d)^\star$ is $\mathbb{R}^d$-valued and the generator is well-defined for each coordinate function, $\mathcal{L}_\lambda f = (\mathcal{L}_\lambda f_1, \ldots, \mathcal{L}_\lambda f_d)^\star$. Furthermore, if F is a $d \times d$ matrix-valued function, $\mathcal{L}_\lambda F$, provided that it is well-defined, denotes the $d \times d$ matrix with ij'th element $\mathcal{L}_\lambda F_{ij}$. The infinitesimal generator notation is useful for expressing the following Lemma 3.2.8.

Lemma 3.2.8. *Suppose that Assumption 3.2.5 holds, and that for some* $k \in \mathbb{N}$*,*

$$f(y, x; \theta) \in C_{2(k+1),0,0}^{pol}(\mathcal{X}^2 \times \Theta)\,.$$

Suppose also that

$$a(y; \theta)\,, b(y; \theta) \in C_{2k,0}^{pol}(\mathcal{X} \times \Theta) \quad \textit{and} \quad c(y, z; \theta) \in C_{2k,0}^{p\text{-}pol}(\mathcal{X} \times \mathbb{R} \times \Theta)\,.$$

Then, for $0 \leq t < t + \Delta \leq t + \Delta_0$ *and* $\lambda \in \Theta$*,*

$$\begin{aligned} &\mathbb{E}_\lambda\left(f(X_{t+\Delta}, X_t; \theta) \mid X_t\right) \\ &= \sum_{i=0}^{k} \frac{\Delta^i}{i!} \mathcal{L}_\lambda^i f(X_t, X_t; \theta) + \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E}_\lambda\left(\mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t\right) du_{k+1} \cdots du_1 \end{aligned}$$

and, furthermore,

$$\int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E}_\lambda\left(\mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t\right) du_{k+1} \cdots du_1 = \Delta^{k+1} R_\lambda(\Delta, X_t; \theta)\,.$$

◇

The first part of Lemma 3.2.8 is effectively a jump-diffusion extension of the expression given by e.g. Florens-Zmirou (1989, Lemma 1) for continuous diffusions. Formula (13) of Masuda (2011) demonstrates a similar expansion for stochastic processes with jumps within his setup. A proof of Lemma 3.2.8 is given in Appendix 3.A.4.

Aside from its application in technial proofs, Lemma 3.2.8 is, together with Assumption 3.2.6.(i), key to proving Lemma 3.2.9, which reveals two important properties of the approximate martingale estimating functions. Lemma 3.2.9 is very similar to Lemma 2.3 of Sørensen (2010), to which we refer for further details on the proof.

Lemma 3.2.9. *Suppose that Assumptions 3.2.5 and 3.2.6 hold. Then, for all $x \in \mathcal{X}$ and $\theta \in \Theta$,*

$$g(0, x, x; \theta) = 0 \quad \text{and} \quad g^{(1)}(x, x; \theta) = -\mathcal{L}_\theta \left(g(0, \theta)\right)(x, x)\,.$$

◇

In concrete examples, Lemma 3.2.8 is also useful for verifying Assumption 3.2.6.(i), a fundamental property of approximate martingale estimating functions, and, conversely, it can be used to create such estimating functions as well.

Remark 3.2.10. Note, for use in the following, that under Assumptions 3.2.5 and 3.2.6,

$$\begin{aligned}
&\mathcal{L}_\lambda(g(0,\theta))(x,x)\\
&\quad = a(x;\lambda)\partial_y g(0,x,x;\theta) + \tfrac{1}{2}b^2(x;\lambda)\partial_y^2 g(0,x,x;\theta)\\
&\qquad + \int_{\mathbb{R}} g(0, x + c(x,z;\lambda), x;\theta)\,\nu_\lambda(dz)\\
&\mathcal{L}_\lambda(\partial_\theta g(0,\theta))(x,x)\\
&\quad = a(x;\lambda)\partial_y\partial_\theta g(0,x,x;\theta) + \tfrac{1}{2}b^2(x;\lambda)\partial_y^2\partial_\theta g(0,x,x;\theta)\\
&\qquad + \int_{\mathbb{R}} \partial_\theta g(0, x + c(x,z;\lambda), x;\theta)\,\nu_\lambda(dz)\\
&\partial_\theta\mathcal{L}_\theta(g(0,\theta))(x,x)\\
&\quad = \mathcal{L}_\theta(\partial_\theta g(0,\theta))(x,x) + \partial_y g(0,x,x;\theta)\partial_\theta a(x;\theta) + \tfrac{1}{2}\partial_y^2 g(0,x,x;\theta)\partial_\theta b^2(x;\theta)\\
&\qquad + \int_{\mathbb{R}} \partial_y g(0, x + c(x,z;\theta), x;\theta)\partial_\theta c(x,z;\theta)\,\nu_\theta(dz)\\
&\qquad + \int_{\mathbb{R}} g(0, x + c(x,z;\theta), x;\theta)\partial_\theta q(z;\theta)\,\tilde{\nu}(dz)\\
&\mathcal{L}_\lambda(gg^\star(0,\theta))(x,x)\\
&\quad = b^2(x;\lambda)\partial_y g\partial_y g^\star(0,x,x;\theta) + \int_{\mathbb{R}} gg^\star(0, x + c(x,z;\lambda), x;\theta)\,\nu_\lambda(dz)
\end{aligned}$$

for all $x \in \mathcal{X}$ and $\lambda, \theta \in \Theta$, by (3.2.8) and Lemmas 3.2.9 and 3.A.2. ○

3.3 General Existence, Uniqueness & Convergence Theorem

This section contains Theorem 3.3.2, the general theorem on the properties of consistent approximate martingale estimating function-based estimators of θ_0 in the model (3.2.1). Assumption 3.3.1 is the final assumption needed for the theorem. The notation A, B and C corresponds to the notation of Lemma 3.5.1, and is also used in Theorem 3.3.2 and its proof.

Assumption 3.3.1. *The following holds for all $\theta \in \Theta$.*

(i) The $\mathbb{R}^d$-vector

$$A(\lambda;\theta) = \int_X (\mathcal{L}_\theta(g(0;\lambda))(x,x) - \mathcal{L}_\lambda(g(0;\lambda))(x,x))\,\pi_\theta(dx)$$

is non-zero whenever $\lambda \neq \theta$.

(ii) The $d \times d$ matrix

$$B(\theta;\theta) = \int_X (\mathcal{L}_\theta(\partial_\theta g(0;\theta))(x,x) - \partial_\theta \mathcal{L}_\theta(g(0;\theta))(x,x))\,\pi_\theta(dx)$$

is non-singular.

(iii) The symmetric $d \times d$ matrix

$$C(\theta;\theta) = \int_X \mathcal{L}_\theta(gg^\star(0,\theta))(x,x)\,\pi_\theta(dx)$$

is positive definite.

◇

In the following, convergence in distribution, denoted $\xrightarrow{\mathcal{D}}$, is assumed to be under the true probability measure $\mathbb{P}_{\theta_0}$ as $n \to \infty$, unless otherwise mentioned.

Theorem 3.3.2. *Suppose that Assumptions 3.2.5, 3.2.6 and 3.3.1 hold. If Assumption 3.2.6.(i) holds with $R_\theta(t,x) \not\equiv 0$, i.e. if $G_n(\theta)$ is not a martingale estimating function, suppose also that $n\Delta_n^{2\kappa-1} \to 0$ as $n \to \infty$. Then,*

(i) there exists a consistent G_n-estimator $\hat{\theta}_n$. Choose any compact, convex set $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$, where $\operatorname{int} K$ denotes the interior of K. Then, $\hat{\theta}_n$ is eventually unique in K, in the sense that for any G_n-estimator $\tilde{\theta}_n$ with $\mathbb{P}_{\theta_0}(\tilde{\theta}_n \in K) \to 1$ as $n \to \infty$, it holds that $\mathbb{P}_{\theta_0}(\hat{\theta}_n \neq \tilde{\theta}_n) \to 0$ as $n \to \infty$.

(ii) for any consistent G_n-estimator $\hat{\theta}_n$, it holds that

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, V(\theta_0)),$$

where

$$V(\theta_0) = B(\theta_0;\theta_0)^{-1} C(\theta_0;\theta_0)(B(\theta_0;\theta_0)^\star)^{-1}$$

is positive definite, and

$$\begin{aligned} B(\theta_0;\theta_0) &= \int_X \left(\mathcal{L}_{\theta_0}(\partial_\theta g(0;\theta_0))(x,x) - \partial_\theta \mathcal{L}_\theta(g(0;\theta))(x,x)|_{\theta=\theta_0}\right)\pi_{\theta_0}(dx), \\ C(\theta_0;\theta_0) &= \int_X \mathcal{L}_{\theta_0}(gg^\star(0,\theta_0))(x,x)\,\pi_{\theta_0}(dx). \end{aligned} \tag{3.3.1}$$

(iii) *for any consistent G_n-estimator $\hat{\theta}_n$,*

$$\widehat{V}_n = n\Delta_n \left(\sum_{i=1}^n \partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) \right)^{-1} \left(\sum_{i=1}^n g g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) \right) \times \left(\sum_{i=1}^n \partial_\theta g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) \right)^{-1}$$

is a consistent estimator of $V(\theta_0)$, so

$$\sqrt{n\Delta_n}\, \widehat{V}_n^{-1/2}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, I_d),$$

where $\widehat{V}_n^{1/2}$ is the unique, positive semidefinite square root of $\widehat{V}_n$ and I_d is the $d \times d$ identity matrix.

◇

While the limit distribution in Theorem 3.3.2.(ii) depends on the unknown parameter θ_0, Theorem 3.3.2.(iii) yields a more practically applicable result. The proof of Theorem 3.3.2 is given in Section 3.5.2.

The stochastic differential equation (3.3.2) and the estimating function used in the following Example 3.3.3 correspond to the ones considered by Masuda (2011), but incorporated into the fully parametric framework of this paper. The asymptotic result (3.3.4) is in accordance with Masuda (2011, Theorem 3.4). Similarly, Theorem 3.3.2.(ii) is comparable to Masuda (2013, Theorem 2.9) (in the case of univariate diffusions), when the Gaussian quasi-likelihood estimator of Masuda is interpreted as an approximate martingale estimating function.

Example 3.3.3. Let the stochastic differential equation

$$dX_t = \tilde{a}(X_t; \alpha)\, dt + \tilde{b}(X_t; \beta)\sigma\, dW_t + \int_{\mathbb{R}} \tilde{b}(X_{t-}; \beta) z\, (N - \mu)(dt, dz) \tag{3.3.2}$$

of the form (3.2.6) be given. The drift parameter α, and the diffusion-jump parameter β are the unknown parameters to be estimated. For simplicity, let $\alpha \in A \subseteq \mathbb{R}$ and $\beta \in B \subseteq \mathbb{R}$ so that $d = 2$ (the results generalise to $d \in \mathbb{N}$ as well). Put $\theta = (\alpha, \beta)^\star$ and $\Theta = A \times B$, and suppose that Assumption 3.2.5 holds. Furthermore, suppose that $\sigma^2 + \gamma_2 = 1$, where γ_k denotes the kth moment of the Lévy measure ν (which does not depend on β).

By Lemma 3.2.8,

$$\mathbb{E}_\theta(X_{t+\Delta} \mid X_t) = X_t + \Delta\, \tilde{a}(X_t; \alpha) + \Delta^2 R_\theta(\Delta, X_t)$$
$$\mathbb{E}_\theta((X_{t+\Delta} - X_t)^2 \mid X_t) = \Delta\, \tilde{b}^2(X_t; \beta) + \Delta^2 R_\theta(\Delta, X_t)$$

for $\theta \in \Theta$ and $0 \le t < t + \Delta \le t + \Delta_0$, so, under suitable conditions on the functions $m_1(x; \theta)$ and $m_2(x; \theta)$,

$$g(t, y, x; \theta) = \begin{pmatrix} m_1(x; \theta)\, (y - x - t\tilde{a}(x; \alpha)) \\ m_2(x; \theta) \left((y - x - t\tilde{a}(x; \alpha))^2 - t\tilde{b}^2(x; \beta) \right) \end{pmatrix} \tag{3.3.3}$$

satisfies Assumption 3.2.6 with $\kappa = 2$.

Suppose also that Assumption 3.3.1 holds, and that $n\Delta_n^3 \to 0$ as $n \to \infty$. Then, by Theorem 3.3.2.(ii), for any consistent G_n-estimator $\hat{\theta}_n$, based on the approximate martingale estimating function $G_n(\theta)$ given by (3.3.3),

$$\sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_2(0, V(\theta_0)) \tag{3.3.4}$$

where $V(\theta_0) = B(\theta_0;\theta_0)^{-1}C(\theta_0;\theta_0)(B(\theta_0;\theta_0)^\star)^{-1}$ with

$$B(\theta_0;\theta_0) = -\int_{\mathcal{X}} \begin{pmatrix} m_1(x;\theta_0)\partial_\alpha \tilde{a}(x;\alpha_0) & 0 \\ 0 & m_2(x;\theta_0)\partial_\beta \tilde{b}^2(x;\beta_0) \end{pmatrix} \pi_{\theta_0}(dx)$$

and

$$C(\theta_0;\theta_0) = \int_{\mathcal{X}} \begin{pmatrix} m_1^2(x;\theta_0)\tilde{b}^2(x;\beta_0) & m_1 m_2(x;\theta_0)\tilde{b}^3(x;\theta_0)\gamma_3 \\ m_2 m_1(x;\theta_0)\tilde{b}^3(x;\theta_0)\gamma_3 & m_2^2(x;\theta_0)\tilde{b}^4(x;\theta_0)\gamma_4 \end{pmatrix} \pi_{\theta_0}(dx)\,.$$

∘

3.4 Rate Optimality and Efficiency

In this section, we approach the challenge of finding rate optimal and efficient estimators in sub-models of (3.2.1). In Section 3.4.1, we present a definition of rate optimality and efficiency. In Section 3.4.2, we propose and motive a conjecture on when a consistent estimator in the type of sub-model under consideration is rate optimal and efficient. In Sections 3.4.3 and 3.4.4, conditions are given on the approximate martingale estimating functions, which ensure rate optimality and efficiency of G_n-estimators in three specific types of sub-models. Section 3.4.5 contains a discussion of the challenge of finding rate optimal and efficient approximate martingale estimating functions, and includes suggestions for future research.

Suppose in the following that $A \subseteq \mathbb{R}^{d_1}$ and $B \subseteq \mathbb{R}^{d_2}$, and consider the stochastic differential equation

$$dX_t = a(X_t;\alpha)\,dt + b(X_t;\beta)\,dW_t + \int_{\mathbb{R}} c(X_{t-}, z;\alpha) N^\alpha(dt,dz)\,, \quad X_0 = U, \tag{3.4.1}$$

for $\alpha \in A$ and $\beta \in B$. The parameters α and β are referred to as the *drift-jump* and *diffusion* parameters respectively. The Poisson random measure $N^\alpha(dt,dz)$ has intensity $\mu_\alpha(dt,dz) = \nu_\alpha(dz)\,dt$, and ν_α has density $q(z;\alpha) = \xi(\alpha)p(z;\alpha)$ with respect to a σ-finite measure $\tilde{\nu}$, where $p(z;\alpha)$ is a probability density. Let $\theta^\star = (\alpha^\star, \beta^\star)$ and $\Theta = A \times B$. For the sake of simplicity, the following assumption is introduced.

Assumption 3.4.1. *Let $c_{x,\alpha}(z) = c(x,z;\alpha)$. For all $x \in \mathcal{X}$ and $\theta \in \Theta$, one of the two following cases (a) or (b) is applicable:*

(a) *The dominating measure $\tilde{\nu}$ is Lebesgue measure. The set*

$$\mathcal{W}(x) = c_{x,\alpha}(\mathbb{R}) = \{w \in \mathbb{R} \mid \text{there exists } z \in \mathbb{R} \text{ with } c_{x,\alpha}(z) = w\}$$

is open and does not depend on α. The mapping $z \mapsto c_{x,\alpha}(z)$ is bijective with a continuously differentiable inverse $w \mapsto c_{x,\alpha}^{-1}(w)$. In this case, let

$$\varphi(x, w; \alpha) = q(c_{x,\alpha}^{-1}(w); \alpha)|\partial_w c_{x,\alpha}^{-1}(w)|\,, \quad w \in \mathcal{W}(x)$$

be the transformation of the Lévy density by $z \mapsto c_{x,\alpha}(z)$, and let η_x denote Lebesgue measure on $\mathcal{W}(x)$.

(b) *The dominating measure $\tilde{\nu}$ is the counting measure on an at most countable set $Q \subset \mathbb{R}$, and $c_{x,\alpha}(z) = c_x(z)$ for all $z \in Q$. In this case, put*

$$\mathcal{W}(x) = c_x(Q) = \{w \in \mathbb{R} \mid \text{there exists } z \in Q \text{ with } c_x(z) = w\}$$

and

$$\varphi(x, w; \alpha) = \sum_{z \in c_x^{-1}(\{w\})} q(z; \alpha)\,,$$

and let η_x denote the counting measure on $\mathcal{W}(x)$.

In both cases, it is assumed that the interchange of differentiation and integration

$$\partial_\theta \left(\int_{\mathcal{W}(x)} g(0, x + w, x; \theta)\varphi(x, w; \alpha)\, \eta_x(dw) \right) = \int_{\mathcal{W}(x)} \partial_\theta (g(0, x + w, x; \theta)\varphi(x, w; \alpha))\, \eta_x(dw)$$

is allowed for all $x \in \mathcal{X}$. $\diamond$

3.4.1 Definitions and Local Asymptotic Normality

When drawing inference on parameters, it is obviously of interest to use the best available estimator. What "best" means, however, is subject to interpretation. For example, estimators deemed to be optimal by theoretical considerations might be computationally infeasible in practice. Nonetheless, here, the optimality of the estimators in question is considered purely from a mathematical perspective. Their practical feasibility lies outside the domain of this paper.

Let $\hat{\theta}_T$ denote a consistent estimator of the d-dimensional parameter θ_0, which is based on observations of $\mathbf{X}$ sampled according to some sampling scheme depending on T, with $T \to \infty$. (In this paper, observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$ at each stage $T = n$, with $t_i^n = i\Delta_n$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$ as $n \to \infty$, unless otherwise mentioned.) We suggest the following definition of rate optimality and efficiency.

Definition 3.4.2. Consider the expression

$$\delta_T(\hat{\theta}_T - \theta_0) \xrightarrow{\mathcal{D}} Z \tag{3.4.2}$$

as $T \to \infty$. Here δ_T denotes a $d \times d$ diagonal matrix with strictly positive entries $(\delta_T)_{jj} \to \infty$ for $j = 1, \ldots, d$, and Z is a zero-mean d-dimensional random vector, with positive definite covariance matrix $\mathcal{V}(\theta_0)$. To the extent that is it possible to derive an asymptotic result of this type for $\hat{\theta}_T$, it is preferred that

(i) the rate of convergence, δ_T, is as fast as possible. If a fastest possible rate $\delta_{0,T}$ has been shown to exist for $\mathbf{X}$ and the sampling scheme considered, and (3.4.2) holds with $(\delta_T)_{jj}/(\delta_{0,T})_{jj} = O(1)$ as $T \to \infty$ for $j = 1, \ldots, d$, the estimator $\hat{\theta}_T$ is said to be *rate optimal*.

(ii) the asymptotic variance $\mathcal{V}(\theta_0)$ is as small as possible. Suppose that $\hat{\theta}_T$ is rate optimal for a specific sampling scheme, and that a smallest possible asymptotic covariance matrix $\mathcal{V}_0(\theta_0)$ has been established in the setup in question, in the sense of partial ordering of positive semidefinite matrices. Then $\hat{\theta}_T$ is said to be *efficient* if (3.4.2) holds with ($\delta_T = \delta_{0,T}$ and) $\mathcal{V}(\theta_0) = \mathcal{V}_0(\theta_0)$.

$\diamond$

Let $\mathcal{G}_T$ be the σ-algebra generated by the observations up to stage T, and let $\mathbb{P}_\theta^T$ denote the restriction of $\mathbb{P}_\theta$ to $\mathcal{G}_T$. Define the likelihood ratios $Q_T(\lambda; \theta) = \log(d\mathbb{P}_\lambda^T / d\mathbb{P}_\theta^T)$, which are supposed to exist for all T, and let $\delta_{0,T}$ be a sequence of invertible, diagonal, $d \times d$ matrices with each entry of $\delta_{0,T}^{-1}$ going to 0 as $T \to \infty$.

Definition 3.4.3. The model $(\Omega, \mathcal{F}, (\mathbb{P}_\theta)_{\theta\in\Theta})$ for $\mathbf{X}$ is said to be *locally asymptotically normal (LAN)* at $\theta_0 \in \Theta$ with *rate* $\delta_{0,T}$ and *asymptotic Fisher information* $\mathcal{I}(\theta_0)$ (under the specified sampling scheme), if the following local asymptotic normality property holds. For all $u \in \mathbb{R}^d$,

$$Q_T\left(\theta_0 + \delta_{0,T}^{-1}u; \theta_0\right) - u^\star S_T(\theta_0) + \tfrac{1}{2}u^\star \mathcal{I}(\theta_0)u \xrightarrow{\mathcal{P}} 0$$

as $T \to \infty$, for some non-random, positive definite $d \times d$ matrix $\mathcal{I}(\theta_0)$, and a sequence $S_T(\theta_0)$ of d-dimensional, $\mathcal{G}_T$-measurable random vectors with $S_T(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, \mathcal{I}(\theta_0))$ as $T \to \infty$. $\diamond$

For more about LAN, see e.g. van der Vaart (2002), who gives a structured overview of Lucien Le Cam's contributions to theoretical statistics (with references in Le Cam (2002)), or Jacod (2010); Le Cam and Yang (2000).

It is seen from the theorem of Hájek (1970), and by Ibragimov and Has'minskii (1981, pp. 152-153, see also Theorem 9.1), that if the statistical model for $\mathbf{X}$ satisfies the local asymptotic normality property under a specified sampling scheme, with rate $\delta_{0,T}$ and asymptotic Fisher information $\mathcal{I}(\theta_0)$, then $\hat{\theta}_T$ is efficient in the sense of Definition 3.4.2.(ii), if (3.4.2) holds with $\delta_T = \delta_{0,T}$ and $\mathcal{V}(\theta_0) = \mathcal{I}(\theta_0)^{-1}$. Thus, Definition 3.4.2 is in accordance with the usual notion of (rate optimality and) efficiency within the framework of local asymptotic normality, see e.g. Jacod (2010, Section 3.1) or Ibragimov and Has'minskii (1981, Definition 11.1).

3.4.2 Conjecture on Rate Optimality and Efficiency

In this section, we first propose our conjecture on rate optimality and efficiency within models of the form (3.4.1). Subsequently, we motivate the conjecture using relevant results from the statistical literature.

Conjecture 3.4.4. *Let* $\mathbf{X}$ *be ergodic, and the unique, strong, càdlàg solution to the stochastic differential equation (3.4.1). Let* $\hat{\theta}_n$ *with* $\hat{\theta}_n^\star = (\hat{\alpha}_n^\star, \hat{\beta}_n^\star)$ *denote a consistent estimator of* θ_0 *based on discrete observations of* $\mathbf{X}$*, sampled at times* $t_i^n = i\Delta_n$*, with* $\Delta_n \to 0$ *and* $n\Delta_n \to \infty$ *as* $n \to \infty$*. Under suitable regularity conditions, and under Assumption 3.4.1,* $\hat{\theta}_n$ *is (rate optimal and) efficient if*

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, \mathcal{V}_0(\theta_0)),$$

where $\mathcal{V}_0(\theta_0)$ *is the (well-defined) inverse of*

$$\mathcal{I}(\theta_0) = \begin{pmatrix} \mathcal{I}_1(\theta_0) & 0 \\ 0 & \mathcal{I}_2(\theta_0) \end{pmatrix},$$

with

$$\mathcal{I}_1(\theta_0) = \int_X \left(\frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x;\beta_0)} + \int_{\mathcal{W}(x)} \frac{\partial_\alpha \varphi(x,w;\alpha_0)^\star \partial_\alpha \varphi(x,w;\alpha_0)}{\varphi(x,w;\alpha_0)} \eta_x(dw) \right) \pi_{\theta_0}(dx)$$

$$\mathcal{I}_2(\theta_0) = \frac{1}{2} \int_X \frac{\partial_\beta b^2(x;\beta_0)^\star \partial_\beta b^2(x;\beta_0)}{b^4(x;\beta_0)} \pi_{\theta_0}(dx).$$

◇

The rest of this section contains some very short summaries of results from the literature, which are used to motivate Conjecture 3.4.4. It is important to note the following limitations, which are imposed in order to keep the discussion as concise as possible:

First, the results quoted from the literature are often presented in a much less general version than what was actually proven in the referenced papers. Results for processes which may be, for example, multivariate, not necessarily ergodic and/or which are permitted to have jumps of infinite activity, are all tailored to fit the more simple framework of our conjecture.

Secondly, no further regularity assumptions are stated in detail. It is understood that each of the quoted results holds under technical conditions stated in its article of origin, and that our conjecture is an informal extrapolation on the basis of these findings.

Let $\widehat{\mathbf{X}}_T$ denote continuous-time observations of the full sample path of $\mathbf{X}$ over the interval $[0,T]$ for $T > 0$, and let $\widehat{X}_n$ denote discrete observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$ of $\mathbf{X}$ sampled as described in the conjecture.

In the case of continuous diffusions, i.e. when $c(x,z;\alpha) \equiv 0$, there exist quite general local asymptotic normality results for the scheme $\widehat{X}_n$. It was shown by Gobet (2002, Theorem

4.1) that the local asymptotic normality property is satisfied with rate $\sqrt{n\Delta_n}$ for α_0, $\sqrt{n}$ for β_0 and asymptotic Fisher information

$$\mathcal{I}(\theta_0) = \begin{pmatrix} \mathcal{I}_1(\theta_0) & 0 \\ 0 & \mathcal{I}_2(\theta_0) \end{pmatrix},$$

where

$$\mathcal{I}_1(\theta_0) = \int_{\mathcal{X}} \frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x;\beta_0)} \pi_{\theta_0}(dx)$$
$$\mathcal{I}_2(\theta_0) = \frac{1}{2}\int_{\mathcal{X}} \frac{\partial_\beta b^2(x;\beta_0)^\star \partial_\beta b^2(x;\beta_0)}{b^4(x;\beta_0)} \pi_{\theta_0}(dx)\,.$$

Sørensen (1991) developed likelihood methods with the purpose of drawing $\widehat{\mathbf{X}}_T$-based inference on the drift-jump parameter α_0 under the assumption that $b(x;\beta) \equiv b(x)$, i.e. that β_0 is known.[4] In case (a) of Assumption 3.4.1, it is seen from formulas (3.4), (3.6) and Corollary 3.3 of Sørensen (1991) that the maximum likelihood estimators $\hat{\alpha}_T$ satisfy that

$$\sqrt{T}(\hat{\alpha}_T - \alpha_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, \mathcal{I}(\alpha_0)^{-1})$$

as $T \to \infty$, where

$$\mathcal{I}(\alpha_0) = \int_{\mathcal{X}} \left(\frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x)} + \int_{\mathcal{W}(x)} \frac{\partial_\alpha \varphi(x,w;\alpha_0)^\star \partial_\alpha \varphi(x,w;\alpha_0)}{\varphi(x,w;\alpha_0)} dw \right) \pi_{\theta_0}(dx)\,. \tag{3.4.3}$$

In the article of Shimizu and Yoshida (2006), a contrast-type estimator $\hat{\theta}_n$ was derived based on $\widehat{X}_n$. In case (a) of Assumption 3.4.1, it is seen from their Theorem 2.1 that if $n\Delta_n^2 \to \infty$, then

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, \mathcal{I}(\theta_0)^{-1})\,,$$

where

$$\mathcal{I}(\theta_0) = \begin{pmatrix} \mathcal{I}_1(\theta_0) & 0 \\ 0 & \mathcal{I}_2(\theta_0) \end{pmatrix}$$

with

$$\mathcal{I}_1(\theta_0) = \int_{\mathcal{X}} \left(\frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x;\beta_0)} + \int_{\mathcal{W}(x)} \frac{\partial_\alpha \varphi(x,w;\alpha_0)^\star \partial_\alpha \varphi(x,w;\alpha_0)}{\varphi(x,w;\alpha_0)} dw \right) \pi_{\theta_0}(dx) \tag{3.4.4}$$

$$\mathcal{I}_2(\theta_0) = \frac{1}{2}\int_{\mathcal{X}} \frac{\partial_\beta b^2(x;\beta_0)^\star \partial_\beta b^2(x;\beta_0)}{b^4(x;\beta_0)} \pi_{\theta_0}(dx)\,.$$

[4] Under this sampling scheme, for all $T > 0$, $\mathbb{P}_\theta^T$ and $\mathbb{P}_{\theta'}^T$ are singular for $\beta \neq \beta'$, making likelihood inference impossible.

Conjecture 3.4.4 is motivated by the following: Suppose case (a) of Assumption 3.4.1. By Shimizu and Yoshida (2006) and Sørensen (1991), it is possible to estimate the drift-jump parameter α on the basis of $\widehat{X}_n$, at the same rate and with the same asymptotic variance as when using maximum likelihood estimation to estimate the parameter using $\widehat{\mathbf{X}}_T$. In our opinion, this maximum likelihood estimator is likely to make optimal use of the information contained in each continuously-observed sample path, so no estimators based on $\widehat{X}_n$ are expected to be able to perform better asymptotically. An argument along these lines also led Shimizu and Yoshida to conclude that their contrast function is efficient for the drift-jump parameter.

Furthermore, by Shimizu and Yoshida (2006) and Gobet (2002), using $\widehat{X}_n$, it is possible to estimate the diffusion parameter β at the same rate and with the same asymptotic variance as when estimating the diffusion parameter efficiently in the corresponding model without jumps. It is our belief that estimation of the diffusion parameter in the latter model should be easier, why estimators obtained for the jump-diffusion model are not expected to be able to perform better asymptotically than efficient estimators pertaining to the corresponding continuous model.

Assume as well, for a moment, that $c(x,z;\alpha) = c(x,z)$ and $q(z;\alpha) = q(z)$. Then (3.4.4) reduces to

$$\mathcal{I}_1(\theta_0) = \int_X \frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x;\beta_0)}\, \pi_{\theta_0}(dx)\,,$$

as does (3.4.3), but with $b(x;\beta) \equiv b(x)$. Consequently, by the result of Shimizu and Yoshida (2006), it is possible to estimate the drift parameter α on the basis of $\widehat{X}_n$, with the same rate and asymptotic variance as in the two following cases: When using maximum likelihood estimation for observations $\widehat{\mathbf{X}}_T$, assuming that there are no unknown diffusion parameters (Sørensen, 1991), and when estimating the drift parameter efficiently in the corresponding model without jumps, also using observations of the type $\widehat{X}_n$ (Gobet, 2002). As in the previous situations, there is no reason to believe that this result can be improved upon in the presence of jumps.

There also exist several local asymptotic normality results in the literature, which are useful to include in the discussion. Suppose, still, that case (a) of Assumption 3.4.1 is applicable, and that

$$dX_t = a(x;\alpha)\,dt + b(X_t)\,dW_t + \int_{\mathbb{R}} z\, N^\alpha(dt,dz)$$

with $\nu_\alpha(dz) = q(z;\alpha)\,dz$. For this model, when $n\Delta_n^2 \to 0$ as $n \to \infty$, Becheri et al. (2014, Propositions 2.1 & 3.1) established the local asymptotic normality property for $\mathbf{X}$ with rate $\sqrt{n\Delta_n}$ for α_0 and asymptotic Fisher information

$$\mathcal{I}(\alpha_0) = \int_X \left(\frac{\partial_\alpha a(x;\alpha_0)^\star \partial_\alpha a(x;\alpha_0)}{b^2(x)} + \int_{\mathbb{R}} \frac{\partial_\alpha q(z;\alpha_0)^\star \partial_\alpha q(z;\alpha_0)}{q(z;\alpha_0)}\, dz \right) \pi_{\alpha_0}(dx) \qquad (3.4.5)$$

within the framework of (and in accordance with) our Conjecture 3.4.4.[5]

[5]The matrix in Assumption 5 of Becheri et al. (2014) may be rewritten to yield (3.4.5).

Suppose now that either case (a) or (b) of Assumption 3.4.1 applies. In the model

$$dX_t = a(x;\alpha)\,dt + b(X_t)\,dW_t + \int_{\mathbb{R}} c(X_{t-},z)\,N(dt,dz)\,,$$

with a one-dimensional drift parameter α, Kohatsu-Higa et al. (2015, Theorem 2.2) showed that $\mathbf{X}$ is locally asymptotically normal with rate $\sqrt{n\Delta_n}$ for α_0 and asymptotic Fisher information

$$\mathcal{I}(\alpha_0) = \int_{\mathcal{X}} \frac{\partial_\alpha a(x;\alpha_0)^2}{b^2(x)}\,\pi_{\alpha_0}(dx)\,,$$

as conjectured above.

Finally, Kohatsu-Higa et al. (2014) considered the model

$$dX_t = (\alpha-\gamma)\,dt + \beta\,dW_t + \int_{\mathbb{R}} z\,N(dt,dz)$$

with $\nu_\gamma(dz) = \gamma\varepsilon_1(dz)$, where ε_1 is the degenerate probability measure with point mass in 1, and the unknown parameter $\theta^\star = (\alpha,\gamma,\beta)$ is three-dimensional. This is an example of case (b) of Assumption 3.4.1. They showed that the model is locally asymptotically normal with rate $\sqrt{n\Delta_n}$ for $(\alpha_0,\gamma_0)^\star$, $\sqrt{n}$ for β_0 and asymptotic Fisher information

$$\mathcal{I}(\theta_0) = \frac{1}{{\beta_0}^2}\begin{pmatrix} 1 & -1 & 0 \\ -1 & (\gamma_0+{\beta_0}^2)/\gamma_0 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

This result is also in accordance with Conjecture 3.4.4.

Further extrapolation on the above leads us to believe that our conjecture holds, under suitable regularity conditions, in each of the two separate cases described in Assumption 3.4.1.

Remark 3.4.5. For use in the following, see that under Assumptions 3.2.5, 3.2.6 and 3.4.1, using Remark 3.2.10, (3.3.1) may be rewritten as

$$\begin{aligned} &B(\theta_0;\theta_0) \\ &\quad= -\int_{\mathcal{X}} \Big(\partial_y g(0,x,x;\theta_0)\partial_\theta a(x;\alpha_0) + \tfrac{1}{2}\partial_y^2 g(0,x,x;\theta_0)\partial_\theta b^2(x;\beta_0)\Big)\,\pi_{\theta_0}(dx) \\ &\qquad- \int_{\mathcal{X}}\int_{\mathcal{W}(x)} g(0,x+w,x;\theta_0)\partial_\theta\varphi(x,w;\alpha_0)\,\eta_x(dw)\,\pi_{\theta_0}(dx)\,, \end{aligned}$$

and

$$\begin{aligned} &C(\theta_0;\theta_0) \\ &\quad= \int_{\mathcal{X}} b^2(x;\beta_0)\partial_y g\partial_y g^\star(0,x,x;\theta_0)\,\pi_{\theta_0}(dx) \\ &\qquad+ \int_{\mathcal{X}}\int_{\mathcal{W}(x)} gg^\star(0,x+w,x;\theta_0)\varphi(x,w;\alpha_0)\,\eta_x(dw)\,\pi_{\theta_0}(dx)\,. \end{aligned}$$

∘

3.4.3 General Drift-Jump Parameter

In this section, a submodel of (3.4.1) with (only) a d-dimensional drift-jump parameter α is considered, that is,

$$dX_t = a(X_t;\alpha)\,dt + b(X_t)\,dW_t + \int_{\mathbb{R}} c(X_{t-},z;\alpha)N^{\alpha}(dt,dz)\,, \quad X_0 = U\,, \tag{3.4.6}$$

with $\alpha \in A$, where $\Theta = A$ is a non-empty, open subset of $\mathbb{R}^d$. According to Conjecture 3.4.4, Theorem 3.3.2 already yields rate optimal estimators of the parameter. In order to ensure efficiency, the following (sufficient) condition is imposed.

Condition 3.4.6 (For use in conjunction with the notation of Assumption 3.4.1)**.** *For each $\alpha \in A$, there exists an invertible $d \times d$ matrix K_α such that for all $x \in \mathcal{X}$,*

$$\partial_y g(0,x,x;\alpha) = K_\alpha \frac{\partial_\alpha a(x;\alpha)^\star}{b^2(x)} \quad \textit{and} \quad g(0,x+w,x;\alpha) = K_\alpha \frac{\partial_\alpha \varphi(x,w;\alpha)^\star}{\varphi(x,w;\alpha)}\,,$$

for η_x-almost all $w \in \mathcal{W}(x)$. ◇

Using Remark 3.4.5, Corollary 3.4.7 follows easily.

Corollary 3.4.7. *Suppose that the assumptions of Theorem 3.3.2, as well as Assumption 3.4.1 and Condition 3.4.6 hold, and that more specifically,* **X** *solves a stochastic differential equation of the form (3.4.6). Then, any consistent G_n-estimator $\hat{\alpha}_n$ is efficient.* ◇

The first equation in Condition 3.4.6 corresponds to the condition given by Sørensen (2010, Condition 1.2) for efficiency of drift parameter-estimators in the case of continuous diffusions.

The second equation marks the introduction of a new type of jump-related condition on the function $g(t,y,x;\alpha)$, not seen in the paper of Sørensen, namely conditions on the off-diagonal $y \neq x$ when $t = 0$. Hypothetically, in the limit $\Delta_n \to 0$, the full sample path of **X** is observed, and whenever relevant, $g(0,y,x;\alpha)$ and its derivatives may be thought of as being evaluated in $y = X_t$ and $x = X_{t-}$. For continuous diffusions, $X_t = X_{t-}$ for all t, in which case it seems plausible that no conditions are needed for $y \neq x$ in order to obtain, e.g. efficiency. For jump-diffusions, however, $X_t \neq X_{t-}$ whenever t is a jump time (while $X_t = X_{t-}$ at all other times), so off-diagonal conditions are not surprising.

The essence of Condition 3.4.6 is that in order to estimate α efficiently, the following must be taken into consideration. Very loosely speaking, in the limit $\Delta_n \to 0$, when applied to "continuous parts" of the data, the estimating function should behave like an efficient approximate martingale estimating function for continuous diffusions, whereas when applied to the jumps, the estimating function must correspond to the score function of the jump. In other words, it is not only necessary to be able to distinguish between the continuous and discontinuous parts of the data asymptotically, but a very specific estimating function must be used for the jump-part.

When differentiated with respect to the parameter, thus yielding a score function approximation comparable to our estimating functions, the efficient contrast function proposed by

Shimizu and Yoshida (2006) satisfies the second equation in Condition 3.4.7 (for π_{θ_0}-a.a. x), although it does not satisfy, e.g. the differentiability assumptions of this paper. (See Section 3.4.5 for further comments on this topic.)

3.4.4 One-Dimensional Diffusion Parameter

In this section, two sub-models of (3.4.1) are considered, both with a one-dimensional diffusion parameter. As a supplement to Assumptions 3.2.5 and 3.2.6, Assumption 3.4.8 is introduced, effectively strengthening the former assumptions in order to obtain rate optimality of estimators of the diffusion parameter. Although only utilised with $d_1 = 0, 2$ and $d_2 = 1$ in this section, the assumption is formulated for more general d_1 and d_2, for use in connection with the auxiliary results in Appendix 3.A.

Assumption 3.4.8. *The parameter set $\Theta = A \times B$ is a non-empty subset of $\mathbb{R}^d$, where A and B are open subsets of $\mathbb{R}^{d_1}$ and $\mathbb{R}^{d_2}$ respectively, with $d_2 \geq 1$ and $d = d_1 + d_2$. Let $\alpha \in A$ and $\beta \in B$ with $\theta^\star = (\alpha^\star, \beta^\star)$. The stochastic process $\mathbf{X} = (X_t)_{t\geq 0}$ solves a stochastic differential equation of the form (3.4.1), the coefficients of which satisfy that*

$$a(y;\alpha) \in C^{pol}_{4,2}(\mathcal{X} \times A), \quad b(y;\beta) \in C^{pol}_{4,2}(\mathcal{X} \times B) \quad \text{and} \quad c(y,z;\alpha) \in C^{p\text{-}pol}_{4,2}(\mathcal{X} \times \mathbb{R} \times A) .$$

The function $g(t,y,x;\theta)$, with $g_\alpha = (g_1, \ldots, g_{d_1})^\star$ and $g_\beta = (g_{d_1+1}, \ldots, g_d)^\star$, satisfies that

$$g_j(t,y,x;\theta) \in C^{pol}_{2,6,1,2}((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$$

for $j = 1, \ldots, d$, and allows the expansion

$$g(\Delta,y,x;\theta) = g(0,y,x;\theta) + \Delta g^{(1)}(y,x;\theta) + \tfrac{1}{2}\Delta^2 g^{(2)}(y,x;\theta) + \Delta^3 R(\Delta,y,x;\theta)$$

where $g^{(i)} = (g^{(i)}_1, \ldots, g^{(i)}_d)^\star$, and $g^{(i)}_j(y,x;\theta)$ is the ith partial derivative of $g_j(t,y,x;\theta)$ with respect to t, evaluated in $t = 0$. ◇

For notational convenience in connection with off-diagonal conditions, Defintion 3.4.9 is made use of as well.

Definition 3.4.9. Define, for $m \in \mathbb{N}$, $\mathbf{z}_m = (z_1, \ldots, z_m)^\star \in \mathbb{R}^m$ and the functions $\tau_m : \mathcal{X} \times \mathbb{R}^m \times A \to \mathcal{X}$ by

$$\tau_m(y,\mathbf{z}_m;\alpha) = \tau_{m-1}(y + c(y,z_m;\alpha), \mathbf{z}_{m-1};\alpha)$$

where $\mathbf{z}_0 = ()$ and $\tau_0(y,\mathbf{z}_0;\alpha) = y$, so that, e.g.

$$\begin{aligned}
\tau_1(y,\mathbf{z}_1;\alpha) &= y + c(y,z_1;\alpha) \\
\tau_2(y,\mathbf{z}_2;\alpha) &= y + c(y,z_2;\alpha) + c(y + c(y,z_2;\alpha), z_1;\alpha) .
\end{aligned}$$

◇

Without Drift-Jump Parameter

In this section, a special case of (3.4.1) with $d_1 = 0$ and $d = d_2 = 1$ is considered, that is, $\mathbf{X}$ solves the stochastic differential equation given by

$$dX_t = a(X_t)\,dt + b(X_t;\beta)\,dW_t + \int_{\mathbb{R}} c(X_{t-}, z) N(dt, dz)\,, \quad X_0 = U\,,$$

for $\beta \in B \subseteq \mathbb{R}$. Condition 3.4.10 is the final condition needed in Theorem 3.4.11. The theorem establishes (sufficient) conditions under which the consistent G_n-estimators $\hat{\beta}_n$, originally discussed in Theorem 3.3.2, are rate optimal in a setup with no drift-jump parameter, and the asymptotic variances can be estimated consistently.

Condition 3.4.10 (For use with Assumption 3.4.8)**.** *Suppose that for all $\beta \in B$,*

$$\begin{aligned} g(0, \tau_k(x, \mathbf{z}_k), x;\beta) &= 0\,, \quad k = 1, 2 \\ \partial_y g(0, \tau_k(x, \mathbf{z}_k), x;\beta) &= 0\,, \quad k = 0, 1 \end{aligned}$$

for all $x \in \mathcal{X}$, and $\tilde{\nu}$-almost all $\mathbf{z}_k \in \mathbb{R}^k$, with $\tau_k(x, \mathbf{z}_k)$ defined in Definition 3.4.9. ◇

Theorem 3.4.11. *Suppose that the assumptions of Theorem 3.3.2, as well as Assumption 3.4.8 and Condition 3.4.10 hold (with $d_1 = 0$ and $d_2 = 1$). If Assumption 3.2.6.(i) holds with $R_\theta(t, x) \not\equiv 0$, i.e. if $G_n(\theta)$ is not a martingale estimating function, suppose also that $n\Delta_n^{2(\kappa-1)} \to 0$ as $n \to \infty$. Let*

$$\begin{aligned} B(\beta_0;\beta_0) &= -\int_{\mathcal{X}} \tfrac{1}{2}\partial_\beta b^2(x;\beta_0)\partial_y^2 g(0, x, x;\beta_0)\,\pi_{\beta_0}(dx)\,, \\ D(\beta;\beta) &= \int_{\mathcal{X}} \tfrac{1}{2} b^4(x;\beta)\partial_y^2 g(0, x, x;\beta)^2\,\pi_\beta(dx)\,, \end{aligned}$$

and suppose that $D(\beta;\beta) > 0$ for all $\beta \in B$. Then, for any consistent G_n-estimator $\hat{\beta}_n$,

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V(\beta_0)) \tag{3.4.7}$$

where $V(\beta_0) = B(\beta_0;\beta_0)^{-2} D(\beta_0;\beta_0) > 0$. Furthermore,

$$\widehat{V}_n = n\left(\sum_{i=1}^n \partial_\beta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\beta}_n)\right)^{-2} \sum_{i=1}^n g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\beta}_n)$$

is a consistent estimator of $V(\beta_0)$, so

$$\sqrt{n}\,\widehat{V}_n^{-1/2}(\hat{\beta}_n - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)\,.$$

◇

The proof of Theorem 3.4.11 is given in Section 3.5.2. By Conjecture 3.4.4, the following Condition 3.4.12 ensuring efficiency is obtained.

Condition 3.4.12. *Suppose that for each $\beta \in B$, there exists a constant $K_\beta \in \mathbb{R}\backslash\{0\}$ such that for all $x \in \mathcal{X}$,*

$$\partial_y^2 g(0,x,x;\beta) = K_\beta \frac{\partial_\beta b^2(x;\beta)}{b^4(x;\beta)} \,.$$

◇

Corollary 3.4.13. *Suppose that the assumptions of Theorem 3.4.11, and Assumption 3.4.1 and Condition 3.4.12 hold. Then, any consistent G_n-estimator $\hat{\beta}_n$ is efficient.* ◇

Condition 3.4.10 for rate optimality and consistent estimation of the asymptotic variance of $\hat{\beta}_n$ is significantly more complicated than the corresponding condition of Sørensen (2010, Condition 1.1), which is the second equation of our condition with $k = 0$. As also observed in Section 3.4.3, for jump-diffusions, conditions also appear on the off-diagonal $y \neq x$ of $g(0,y,x;\beta)$ and selected derivatives.

Condition 3.4.10 does suggest that for models with certain jump dynamics (certain combinations of c and $\tilde{\nu}$), rate optimal estimation of the diffusion parameter might not be feasible within the framework of this paper. If, for example, the first equation amounts to the requirement that

$$g(0,y,x;\beta) = 0 \tag{3.4.8}$$

for all $x, y \in \mathcal{X}$, e.g. the non-degeneracy condition on $D(\beta;\beta)$ in Theorem 3.4.11 becomes impossible to satisfy.

When the efficient contrast function of Shimizu and Yoshida (2006) is differentiated with respect to the parameter (and multiplied by Δ_n), the resulting function easily satisfies the first equation in Condition 3.4.10. In fact, it satisfies equation (3.4.8) for all $x, y \in \mathcal{X}$ by the help of an indicator function depending on, among other things, x and y, thus satisfying the rest of Condition 3.4.10 as well. However, as mentioned previously, due to its general non-differentiability, their function cannot readily be adapted to our setup. Also, it does not satisfy the above-mentioned non-degeneracy condition on $D(\beta;\beta)$.

The additional condition for efficiency of the rate optimal estimators of Theorem 3.4.11, Condition 3.4.12, is the same as the one identified by Sørensen (2010) for continuous diffusions.[6]

3.4.4.1 Two-Dimensional Drift-Jump Parameter

This section considers a slightly more general model than the previous section, namely, one which includes both a two-dimensional drift-jump parameter α and a one-dimensional diffusion parameter β. The model is a special case of (3.4.1) with $d_1 = 2$ and $d_2 = 1$ ($d = 3$), i.e.

$$dX_t = a(X_t;\alpha)\,dt + b(X_t;\beta)\,dW_t + \int_{\mathbb{R}} c(X_{t-},z;\alpha)N^\alpha(dt,dz)\,, \quad X_0 = U\,,$$

[6]And the same condition obtained for efficiency in Chapter 2 under a different observation scheme.

for $\alpha \in A \subseteq \mathbb{R}^2$ and $\beta \in B \subseteq \mathbb{R}$, with $\theta^\star = (\alpha^\star, \beta)$ and $\Theta = A \times B$.

The following Condition 3.4.14 is an additional condition for use in Theorem 3.4.15. Within the framework of the current model, this theorem establishes rate optimality of the consistent G_n-estimators $\hat\theta_n$ with $\hat\theta_n^\star = (\hat\alpha_n^\star, \hat\beta_n)$ obtained by Theorem 3.3.2, and ensures that their asymptotic variances may be estimated consistently. The coordinates $\hat\alpha_n$ estimating the drift-jump parameter already converge at the optimal rate by Theorem 3.3.2.(ii), so it is essentially the convergence rate of the coordinate $\hat\beta_n$, which estimates the diffusion parameter, that is improved upon.

Condition 3.4.14 (For use with Assumption 3.4.8)**.** *Suppose that for all $\tilde\alpha \in A$, $\theta \in \Theta$,*

$$\begin{aligned} g_\beta(0, \tau_k(x, \mathbf{z}_k; \tilde\alpha), x; \theta) &= 0\,, \quad k = 1,2,3,4 \\ \partial_y g_\beta(0, \tau_k(x, \mathbf{z}_k; \tilde\alpha), x; \theta) &= 0\,, \quad k = 0,1,2,3 \\ \partial_y^2 \partial_\alpha g_\beta(0, \tau_k(x, \mathbf{z}_k; \tilde\alpha), x; \theta) &= 0\,, \quad k = 0,1 \\ \partial_\alpha g_\beta^{(1)}(\tau_1(x, z_1; \tilde\alpha), x; \theta) &= 0 \end{aligned}$$

for all $x \in \mathcal{X}$ and $\tilde\nu$-almost all $\mathbf{z}_k \in \mathbb{R}^k$, where $\tau_k(x, \mathbf{z}_k; \tilde\alpha)$ is defined in Definition 3.4.9. ◇

Theorem 3.4.15. *Suppose that the assumptions of Theorem 3.3.2, as well as Assumption 3.4.8 and Condition 3.4.14 hold (with $d_1 = 2$ and $d_2 = 1$). If Assumption 3.2.6.(i) holds with $R_\theta(t, x) \not\equiv 0$, i.e. if $G_n(\theta)$ is not a martingale estimating function, suppose also that $n\Delta_n^{2(\kappa-1)} \to 0$ as $n \to \infty$. Let*

$$\begin{aligned} B_1(\theta_0; \theta_0) = &- \int_{\mathcal{X}} \partial_y g_\alpha(0, x, x; \theta_0) \partial_\alpha a(x; \alpha_0)\, \pi_{\theta_0}(dx) \\ &- \int_{\mathcal{X}} \int_{\mathbb{R}} \partial_y g_\alpha(0, x + c(x, z; \alpha_0), x; \theta_0) \partial_\alpha c(x, z; \alpha_0)\, \nu_{\alpha_0}(dz)\, \pi_{\theta_0}(dx) \\ &- \int_{\mathcal{X}} \int_{\mathbb{R}} g_\alpha(0, x + c(x, z; \alpha_0), x; \theta_0) \partial_\alpha q(z; \alpha_0)\, \tilde\nu(dz)\, \pi_{\theta_0}(dx)\,, \\ B_2(\theta_0; \theta_0) = &- \int_{\mathcal{X}} \tfrac{1}{2} \partial_y^2 g_\beta(0, x, x; \theta_0) \partial_\beta b^2(x; \beta_0)\, \pi_{\theta_0}(dx)\,, \\ E_1(\theta; \theta) = &\int_{\mathcal{X}} b^2(x; \beta) \partial_y g_\alpha \partial_y g_\alpha^\star(0, x, x; \theta)\, \pi_\theta(dx) \\ &+ \int_{\mathcal{X}} \int_{\mathbb{R}} g_\alpha g_\alpha^\star(0, x + c(x, z; \alpha), x; \theta)\, \nu_\alpha(dz)\, \pi_\theta(dx)\,, \\ E_2(\theta; \theta) = &\int_{\mathcal{X}} \tfrac{1}{2} b^4(x; \beta) \partial_y^2 g_\beta(0, x, x; \theta)^2\, \pi_\theta(dx)\,, \end{aligned}$$

and assume that $E_1(\theta; \theta)$ is invertible and $E_2(\theta; \theta) \neq 0$ for all $\theta \in \Theta$. Then, for any consistent G_n-estimator $\hat\theta_n$, it holds that

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat\alpha_n - \alpha_0) \\ \sqrt{n}(\hat\beta_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, V(\theta_0)) \tag{3.4.9}$$

where

$$V(\theta_0) = \begin{pmatrix} B_1(\theta_0;\theta_0)^{-1}E_1(\theta_0;\theta_0)(B_1(\theta_0;\theta_0)^\star)^{-1} & 0 \\ 0 & B_2(\theta_0;\theta_0)^{-2}E_2(\theta_0;\theta_0) \end{pmatrix}$$

is positive definite. Furthermore,

$$\widehat{V}_n = \begin{pmatrix} \widehat{V}_{n,1} & 0 \\ 0 & \widehat{V}_{n,2} \end{pmatrix}$$

given by

$$\begin{aligned} \widehat{V}_{n,1} &= n\Delta_n \left(\sum_{i=1}^n \partial_\alpha g_\alpha(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n)\right)^{-1} \left(\sum_{i=1}^n g_\alpha g_\alpha^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n)\right) \\ &\quad \times \left(\sum_{i=1}^n \partial_\alpha g_\alpha^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n)\right)^{-1} \\ \widehat{V}_{n,2} &= n\left(\sum_{i=1}^n \partial_\beta g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n)\right)^{-2} \sum_{i=1}^n g_\beta^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) \end{aligned}$$

is a consistent estimator of $V(\theta_0)$*, so*

$$\widehat{V}_n^{-1/2} \begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, I_3)\,,$$

where $\widehat{V}_n^{1/2}$ *is the unique, positive semi-definite square root of* $\widehat{V}_n$*.* ◇

The proof of Theorem 3.4.15 is given in Section 3.5.2. Making use of Remark 3.4.5, the following additional Condition 3.4.16 is obtained for efficiency.

Condition 3.4.16 (For use in conjunction with Assumption 3.4.1)**.** *For all* $\theta \in \Theta$ *there exists an invertible* 2×2 *matrix* $K_\theta^{(1)}$ *and a non-zero constant* $K_\theta^{(2)}$*, such that for all* $x \in \mathcal{X}$*,*

$$\begin{aligned} \partial_y g_\alpha(0,x,x;\theta) &= K_\theta^{(1)} \frac{\partial_\alpha a(x;\alpha)^\star}{b^2(x;\beta)}\,, \\ \partial_y^2 g_\beta(0,x,x;\theta) &= K_\theta^{(2)} \frac{\partial_\beta b^2(x;\beta)}{b^4(x;\beta)}\,, \\ g_\alpha(0,x+w,x;\theta) &= K_\theta^{(1)} \frac{\partial_\alpha \varphi(x,w;\alpha)^\star}{\varphi(x,w;\alpha)} \end{aligned}$$

for η_x*-almost all* $w \in \mathcal{W}(x)$*.* ◇

Corollary 3.4.17. *Suppose that the assumptions of Theorem 3.4.15, as well as Assumption 3.4.1 and Condition 3.4.16 hold. Then, any consistent* G_n*-estimator* $\hat{\theta}_n$ *is efficient.* ◇

Comparing Condition 3.4.14 to the corresponding Condition 3.4.10 for the model with no drift-jump parameter, the number of conditions used to obtain rate optimality of the estimators $\hat{\beta}_n$ (and ensure that their asymptotic variances may be estimated consistently) is seen to have increased substantially. The general reason is that, loosely put, the larger the dimension d of the parameter is, the more conditions are needed to show uniform convergence in probability with rate optimality of $\hat{\beta}_n$ in mind, especially when the model includes a drift-jump parameter as well.

The additional Condition 3.4.16 for efficiency is well in line with our previously obtained efficiency conditions, Conditions 3.4.6 and 3.4.12.

3.4.5 On the Existence of Efficient Estimating Functions

For continuous diffusions, conditions under which an approximate martingale estimating function is rate optimal and efficient are quite straightforward, and it is easy to find estimating functions which satisfy the conditions. This was concluded by Sørensen (2010) for the current sampling scheme, and in Chapter 2 for fixed-interval asymptotics. The same cannot be said in the presence of jumps.

Conditions 3.4.10 and 3.4.14 were obtained for rate optimality of the estimator of a one-dimensional diffusion parameter, in a model with no drift-jump parameter or a two-dimensional drift-jump parameter respectively. Essentially, the g_β coordinate of g, as well as several of its derivatives, need to vanish at a number of points depending on the jump dynamics of the process, in order to achieve rate optimality. For some stochastic differential equations, it could be difficult, or perhaps even impossible, to find estimating functions which satisfy these rate optimality conditions, as well as the remaining regularity assumptions. For example, Theorems 3.4.11 and 3.4.15 require that $\partial_y^2 g_\beta(0, x, x; \theta)$ does not vanish π_θ-almost surely for any θ, which could very easily conflict with the rate optimality conditions.

Conditions 3.4.6, 3.4.12 and 3.4.16 were the supplementary conditions obtained for the efficiency of rate optimal estimators of the drift-jump and diffusion parameters in three different models. In addition to the usual conditions for efficiency of approximate martingale estimating functions for continuous diffusions (see again Sørensen (2010)[7]), the present conditions include a very specific requirement, tied to the jumps of the process. The essence of this condition is that in the limit $\Delta_n \to 0$, evaluating $g_\alpha(0, y, x; \theta)$ at a jump increment $(y, x) = (X_t, X_{t-})$ with $X_t \neq X_{t-}$ should be the same as evaluating the score function of the the jump. This entails, in particular, that the g_α-coordinates of an efficient estimating function can discriminate, asymptotically, between situations where $X_t = X_{t-}$ and $X_t \neq X_{t-}$, because, when $X_t = X_{t-}$, the function must behave like an efficient estimating function for the drift parameter of the corresponding continuous diffusion.

The assumption that $g(t, y, x; \theta)$ is sufficiently continuously differentiable with respect to, e.g. t, y and x seems to be the main property distancing the approximate martingale estimating functions considered in this paper from the efficient contrast function of Shimizu and Yoshida (2006). Their contrast function contains indicator functions of the form $\mathbf{1}(|y - x| \leq t^\rho)$, and is thus designed to distinguish asymptotically between jumps and "no-jumps". For

[7] And Chapter 2.

multidimensional drift-jump and diffusion parameters α and β, estimation based on their contrast function corresponds to solving the estimating equation $H_n(\theta) = 0$ where

$$H_n(\theta) = \sum_{i=1}^{n} h(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$$

and $h^\star = (h_\alpha^\star, h_\beta^\star)$ with

$$\begin{aligned} h_\alpha(t,y,x;\theta) &= \frac{\partial_\alpha a(x;\alpha)^\star}{b^2(x;\beta)}(y - x - ta(x;\alpha))\mathbf{1}(|y-x| \leq t^\rho) \\ &\quad + \frac{\partial_\alpha \varphi(x, y-x;\alpha)^\star}{\varphi(x,y-x;\alpha)}\mathbf{1}(|y-x| > t^\rho) - t\int_{\mathcal{W}(x)} \partial_\alpha \varphi(x,w;\alpha)^\star\, dw\,, \\ h_\beta(t,y,x;\theta) &= \frac{\partial_\beta b^2(x;\beta)^\star}{b^4(x;\beta)}\Big((y-x-ta(x;\alpha))^2 - tb^2(x;\beta)\Big)\mathbf{1}(|y-x| \leq t^\rho)\,. \end{aligned}$$

(For simplicity, we assume here that the additional truncation function used by Shimizu and Yoshida to ensure integrability is not necessary for the model under consideration.) Disregarding the indicator functions, h_β and the first term in h_α yield approximate martingale estimating functions for the continuous diffusion corresponding to $\mathbf{X}$, obtained by setting $c(x,z;\alpha) \equiv 0$ (this may be checked using Lemma 3.2.8). The remaining terms in h_α give rise to an approximation of the score function of the compound Poisson jump-part of $\mathbf{X}$. It is not immediately obvious whether or not h satisfies the approximate martingale property of Assumption 3.2.6.(i).

However, suppose, for example, that $c(x,z;\alpha) > c_0$ for some real-valued constant $c_0 > 0$. Let the indicator function $\mathbf{1}(|y-x| \leq t^\rho)$ be replaced by a suitable approximation $\psi(t,y,x)$, twice differentiable with respect to y, once differentiable with respect to t, and satisfying that $\psi(0,x,x) = 1$ and $\psi(0,y,x) = 0$ for $y \geq x + c_0$, for all $x \in \mathcal{X}$. Then,

$$\begin{aligned} \tilde{h}_\alpha(t,y,x;\theta) &= \frac{\partial_\alpha a(x;\alpha)^\star}{b^2(x;\beta)}(y-x-ta(x;\alpha))\psi(t,y,x) \\ &\quad + \frac{\partial_\alpha \varphi(x,y-x;\alpha)^\star}{\varphi(x,y-x;\alpha)}(1-\psi(t,y,x)) - t\int_{\mathcal{W}(x)} \partial_\alpha\varphi(x,w;\alpha)^\star\, dw \\ \tilde{h}_\beta(t,y,x;\theta) &= \frac{\partial_\beta b^2(x;\beta)^\star}{b^4(x;\beta)}\Big((y-x-ta(x;\alpha))^2 - tb^2(x;\beta)\Big)\psi(t,y,x)\,. \end{aligned}$$

For example, for a two-dimensional drift-jump parameter, and a one-dimensional diffusion parameter, it may be verified that $\tilde{h}(t,y,x;\theta)$ satisfies the rate optimality and efficiency Conditions 3.4.14 and 3.4.16. Furthermore, supposing for further simplicity that α only enters into the drift coefficient $a(x;\alpha)$, $\tilde{h}_\alpha(t,y,x;\theta)$ reduces to

$$\tilde{h}_\alpha(t,y,x;\theta) = \frac{\partial_\alpha a(x;\alpha)^\star}{b^2(x;\beta)}(y-x-ta(x;\alpha))\psi(t,y,x)\,.$$

It may now be seen that under the additional (and not unreasonable) assumption $\partial_y\psi(0,x,x) = 0$ for $x \in \mathcal{X}$, the equations

$$\begin{aligned} \tilde{h}(0,x,x;\theta) &= 0 \\ \tilde{h}^{(1)}(x,x;\theta) &= -\mathcal{L}_\theta(\tilde{h}(0;\theta))(x,x) \end{aligned}$$

hold for all $x \in \mathcal{X}$ and $\theta \in \Theta$. These identities were established in Lemma 3.2.9 for approximate martingale estimating functions, and are used in many of the proofs in this paper.

It is our hypothesis for future research that under further conditions on ψ (and perhaps more relaxed regularity assumptions than those of this paper) we can find sub-models of (3.4.1), and explicit choices of ψ, for which the functions $\tilde{h}(t, y, x; \theta)$ (or similar approximations to $h(t, y, x; \theta)$) constitute (rate optimal and) efficient, approximate martingale estimating functions. We believe that upon finding a suitable type of ψ-function, other functions resembling $h(t, y, x; \theta)$, with suitable approximate martingale-type components combined with ψ-functions, may be utilised in order to establish a more general class of explicit, efficient approximate martingale estimating functions for jump-diffusions.

3.5 Proofs

Section 3.5.1 states lemmas needed to prove Theorems 3.3.2, 3.4.11 and 3.4.15. These theorems are proven in Section 3.5.2, while the lemmas are proven in Section 3.5.3.

3.5.1 Main Lemmas

The lemmas presented in this section are used, together with results on the existence, uniqueness and convergence of G_n-estimators from Sørensen (2012, Section 1.10), to prove Theorems 3.3.2, 3.4.11 and 3.4.15. For convenience, Theorems 1.58, 1.59 and 1.60 of Sørensen are briefly summarised in Appendix 3.B.2, in a simplified form, tailored to fit the framework of the current paper.

Lemma 3.5.1. *Suppose that Assumptions 3.2.5 and 3.2.6 hold. If Assumption 3.2.6.(i) holds with $R_\theta(t, x) \not\equiv 0$, i.e. if $G_n(\theta)$ is not a martingale estimating function, suppose also that $n\Delta_n^{2\kappa-1} \to 0$ as $n \to \infty$. Let*

$$A(\theta; \theta_0) = \int_{\mathcal{X}} \left(\mathcal{L}_{\theta_0}(g(0; \theta))(x, x) - \mathcal{L}_\theta(g(0; \theta)(x, x)\right) \pi_{\theta_0}(dx)$$
$$B(\theta; \theta_0) = \int_{\mathcal{X}} \left(\mathcal{L}_{\theta_0}(\partial_\theta g(0; \theta))(x, x) - \partial_\theta \mathcal{L}_\theta(g(0; \theta))(x, x)\right) \pi_{\theta_0}(dx)$$
$$C(\theta; \theta_0) = \int_{\mathcal{X}} \mathcal{L}_{\theta_0}(gg^\star(0, \theta))(x, x) \, \pi_{\theta_0}(dx)$$

for $\theta \in \Theta$. Then,

(i) the mappings $\theta \mapsto A(\theta; \theta_0)$, $\theta \mapsto B(\theta; \theta_0)$ and $\theta \mapsto C(\theta; \theta_0)$ are continuous on Θ, with $A(\theta_0; \theta_0) = 0$ and $\partial_\theta A(\theta; \theta_0) = B(\theta; \theta_0)$.

(ii) for all $j, k = 1, \ldots, d$, and all compact, convex sets $K \subseteq \Theta$,

$$\sup_{\theta \in K} \left| \frac{1}{n\Delta_n} \sum_{i=1}^n g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) - A_j(\theta; \theta_0) \right| \xrightarrow{\mathcal{P}} 0 \,,$$

$$\sup_{\theta\in K}\left|\frac{1}{n\Delta_n}\sum_{i=1}^n \partial_{\theta_k} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) - B_{jk}(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0\,. \tag{3.5.1}$$

(iii) for any consistent estimator $\hat{\theta}_n$, it holds that

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\theta}_n) \xrightarrow{\mathcal{P}} B(\theta_0;\theta_0)\,,$$
$$\frac{1}{n\Delta_n}\sum_{i=1}^n gg^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\theta}_n) \xrightarrow{\mathcal{P}} C(\theta_0;\theta_0)\,.$$

(iv) it holds that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, C(\theta_0;\theta_0))\,.$$

◇

Lemma 3.5.2. *Suppose that Assumptions 3.2.5, 3.2.6, and 3.4.8, and Condition 3.4.10 hold (with $d_1 = 0$ and $d_2 = 1$). If Assumption 3.2.6.(i) holds with $R_\theta(t,x) \not\equiv 0$, suppose also that $n\Delta_n^{2(\kappa-1)} \to 0$ as $n \to \infty$. Let*

$$\begin{aligned}D(\beta;\beta_0)\\ &= \int_{\mathcal{X}} \tfrac{1}{2}\left(b^4(x;\beta_0) + \tfrac{1}{2}\left(b^2(x;\beta_0) - b^2(x;\beta)\right)^2\right)\partial_y^2 g(0,x,x;\beta)^2\,\pi_{\beta_0}(dx)\end{aligned}$$

for $\beta \in B$. Then,

(i) for any consistent estimator $\hat{\beta}_n$,

$$\frac{1}{n\Delta_n^2}\sum_{i=1}^n g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\beta}_n) \xrightarrow{\mathcal{P}} D(\beta_0;\beta_0)\,.$$

(ii) it holds that

$$\frac{1}{\sqrt{n}\Delta_n}\sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, D(\beta_0;\beta_0))\,. \tag{3.5.2}$$

◇

Lemma 3.5.3. *Suppose that Assumptions 3.2.5, 3.2.6, and 3.4.8, and Condition 3.4.14 hold (with $d_1 = 2$ and $d_2 = 1$). If Assumption 3.2.6.(i) holds with $R_\theta(t,x) \not\equiv 0$, suppose also that $n\Delta_n^{2(\kappa-1)} \to 0$ as $n \to \infty$. Let*

$$\delta_n = \begin{pmatrix}\sqrt{n\Delta_n} & 0 & 0\\ 0 & \sqrt{n\Delta_n} & 0\\ 0 & 0 & \sqrt{n}\end{pmatrix} \quad \text{and} \quad E(\theta_0;\theta_0) = \begin{pmatrix}E_1(\theta_0;\theta_0) & 0\\ 0 & E_2(\theta_0;\theta_0)\end{pmatrix}$$

with

$$E_1(\theta_0;\theta_0) = \int_{\mathcal{X}} b^2(x;\beta_0)\partial_y g_\alpha \partial_y g_\alpha^\star(0,x,x;\theta_0)\,\pi_{\theta_0}(dx) + \int_{\mathcal{X}}\int_{\mathbb{R}} g_\alpha g_\alpha^\star(0,x+c(x,z;\alpha_0),x;\theta_0)\,\nu_{\alpha_0}(dz)\,\pi_{\theta_0}(dx)\,,$$

$$E_2(\theta_0;\theta_0) = \int_{\mathcal{X}} \tfrac{1}{2} b^4(x;\beta_0)\partial_y^2 g_\beta(0,x,x;\theta_0)^2\,\pi_{\theta_0}(dx)\,.$$

Then,

(i) for $j = 1, 2$, and all compact, convex sets $K \subseteq \Theta$,

$$\sup_{\theta\in K}\left|\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n \partial_{\alpha_j} g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\right| \xrightarrow{\mathcal{P}} 0\,.$$

(ii) for any consistent estimator $\hat\theta_n$,

$$\frac{1}{n\Delta_n^2}\sum_{i=1}^n g_\beta^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat\theta_n) \xrightarrow{\mathcal{P}} E_2(\theta_0;\theta_0)\,.$$

(iii) it holds that

$$\delta_n \frac{1}{n\Delta_n}\sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_3\left(0, E(\theta_0;\theta_0)\right)\,. \tag{3.5.3}$$

◇

3.5.2 Proofs of Main Theorems

This section contains the proofs of Theorems 3.3.2, 3.4.11 and 3.4.15.

Proof of Theorem 3.3.2. Let any compact, convex set $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$ be given, and recall that

$$G_n(\theta) = \frac{1}{n\Delta_n}\sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta).$$

Note that uniform convergence in probability (for θ in compact, convex sets) of a vector or matrix is implied by the corresponding convergence of each of its coordinates.

By Lemma 3.5.1.(i) and (ii), and Assumption 3.3.1.(ii),

$$G_n(\theta_0) \xrightarrow{\mathcal{P}} 0 \quad \text{and} \quad \sup_{\theta\in K}\|\partial_\theta G_n(\theta) - B(\theta;\theta_0)\| \xrightarrow{\mathcal{P}} 0$$

with $B(\theta_0;\theta_0)$ invertible. That is, the estimating function $G_n(\theta)$ satisfies the conditions of Theorem 3.B.2 (Sørensen, 2012, Theorem 1.58).

Next, we show (3.B.1) of Theorem 3.B.3 (Sørensen, 2012, Theorem 1.59). Let $\varepsilon > 0$ be given, and let $\bar{B}_\varepsilon(\theta_0)$ and $B_\varepsilon(\theta_0)$, respectively, denote closed and open balls in $\mathbb{R}^d$ with radius $\varepsilon > 0$, centered at θ_0. The compact set $K \backslash B_\varepsilon(\theta_0)$ does not contain θ_0, and so, by Assumption 3.3.1.(i), $A(\theta, \theta_0) \neq 0$ for all $\theta \in K \backslash B_\varepsilon(\theta_0)$. Then, by the continuity of $\theta \mapsto \|A(\theta, \theta_0)\|$,

$$\inf_{\theta \in K \backslash \bar{B}_\varepsilon(\theta_0)} \|A(\theta, \theta_0)\| \geq \inf_{\theta \in K \backslash B_\varepsilon(\theta_0)} \|A(\theta, \theta_0)\| > 0\,.$$

Now, by Theorem 3.B.3, it follows that for any G_n-estimator $\tilde{\theta}_n$,

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K \backslash \bar{B}_\varepsilon(\theta_0)\right) \to 0 \quad \text{as} \quad n \to \infty \tag{3.5.4}$$

for any $\varepsilon > 0$.

By Theorem 3.B.2, there exists a consistent G_n-estimator $\hat{\theta}_n$ which is eventually unique in the sense that if $\bar{\theta}_n$ is another consistent G_n-estimator, then

$$\mathbb{P}_{\theta_0}\left(\hat{\theta}_n \neq \bar{\theta}_n\right) \to 0 \tag{3.5.5}$$

as $n \to \infty$. Suppose that $\tilde{\theta}_n$ is any G_n-estimator which satisfies that $\mathbb{P}_{\theta_0}(\tilde{\theta}_n \in K) \to 1$ as $n \to \infty$. By (3.5.4), also

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K^c \cup \bar{B}_\varepsilon(\theta_0)\right) \to 1$$

as $n \to \infty$, and it follows that $\tilde{\theta}_n$ is consistent. An application of (3.5.5) completes the proof of Theorem 3.3.2.(i).

By Lemma 3.5.1.(iv),

$$\delta_n G_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, C(\theta_0; \theta_0))\,,$$

where $\delta_n = \sqrt{n\Delta_n}\mathrm{I}_d$ (I_d is the $d \times d$ identity matrix), and the matrix $C(\theta_0; \theta_0)$ is positive definite by Assumption 3.3.1.(iii). Also, note that $\delta_n \partial_\theta G_n(\theta) \delta_n^{-1} = \partial_\theta G_n(\theta)$, so

$$\sup_{\theta \in K} \left\| \delta_n \partial_\theta G_n(\theta) \delta_n^{-1} - B(\theta; \theta_0) \right\| \xrightarrow{\mathcal{P}} 0$$

with $B(\theta_0; \theta_0)$ invertible, as stated previously. Now, Theorem 3.3.2.(ii) follows from Theorem 3.B.4 (Sørensen, 2012, Theorem 1.60), with $V(\theta_0)$ positive definite by Assumption 3.3.1.

Finally, Theorem 3.3.2.(iii) follows from Lemma 3.5.1.(iii) and the continuous mapping theorem. In this connection, it is important to note that $V(\theta_0)$ is non-random, and that taking the unique, positive semi-definite (*principal*) square root of a positive semi-definite real matrix is a continuous transformation. □

Proof of Theorem 3.4.11. Let any compact, convex set $K \subseteq B$ with $\beta_0 \in \operatorname{int} K$ be given, and recall that

$$G_n(\beta) = \frac{1}{n\Delta_n} \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta)\,.$$

As observed in the proof of Theorem 3.3.2, $G_n(\beta)$ satisfies the conditions of Theorem 3.B.2 (Sørensen, 2012, Theorem 1.58). This leaves the remaining conditions of Theorem 3.B.4 (Sørensen, 2012, Theorem 1.60) to be verified. Let $\delta_n = \sqrt{n}$. By Lemma 3.5.2.(ii),

$$\delta_n G_n(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, D(\beta_0;\beta_0)) .$$

Furthermore, $\delta_n \partial_\beta G_n(\beta)\delta_n^{-1} = \partial_\beta G_n(\beta)$, so

$$\sup_{\beta\in K} \left\| \delta_n \partial_\beta G_n(\beta)\delta_n^{-1} - B(\beta;\beta_0) \right\| \xrightarrow{\mathcal{P}} 0$$

continues to hold by (3.5.1), where $B(\beta;\beta_0)$ is as given by Lemma 3.5.1. In particular, $B(\beta_0;\beta_0) \neq 0$. Now, (3.4.7) follows from Theorem 3.B.4. Lemma 3.5.1.(iii) yields

$$\frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_\beta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\beta}_n) \xrightarrow{\mathcal{P}} B(\beta_0;\beta_0) ,$$

which, used together with Lemma 3.5.2.(i) and the continuous mapping theorem, completes the proof. □

Proof of Theorem 3.4.15. Let any compact, convex set $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$ be given. Still, it is seen directly from Lemma 3.5.1 and Assumption 3.3.1 that

$$G_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)$$

satisfies the conditions of Theorem 3.B.2 (Sørensen, 2012, Theorem 1.58). It remains to verify the subsequent conditions of Theorem 3.B.4 (Sørensen, 2012, Theorem 1.60). Let

$$\delta_n = \begin{pmatrix} \sqrt{n\Delta_n} & 0 & 0 \\ 0 & \sqrt{n\Delta_n} & 0 \\ 0 & 0 & \sqrt{n} \end{pmatrix}$$

and see that by Lemma 3.5.3.(iii),

$$\delta_n G_n(\theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, E(\theta_0;\theta_0)) .$$

Observe that

$$\delta_n \partial_\theta G_n(\theta) \delta_n^{-1} = \begin{pmatrix} \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_\alpha g_\alpha(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) & \frac{1}{n\Delta_n^{1/2}} \sum_{i=1}^{n} \partial_\beta g_\alpha(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \\ \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^{n} \partial_\alpha g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) & \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_\beta g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \end{pmatrix},$$

and recall that $g_\alpha = (g_1, g_2)^\star$ and $g_\beta = g_3$, and that $\theta^\star = (\alpha^\star, \beta)$ with $\alpha^\star = (\theta_1, \theta_2)$ and $\beta = \theta_2$. Let $B(\theta;\theta_0) = (B_{jk}(\theta;\theta_0))_{j,k=1,2,3}$ be as given in Lemma 3.5.1. By (3.5.1), for $j, k = 1, 2$,

$$\begin{aligned} &\sup_{\theta\in K} \left| \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_{\theta_k} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) - B_{jk}(\theta;\theta_0) \right| \xrightarrow{\mathcal{P}} 0 \\ &\sup_{\theta\in K} \left| \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_{\theta_3} g_3(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) - B_{33}(\theta;\theta_0) \right| \xrightarrow{\mathcal{P}} 0 . \end{aligned} \tag{3.5.6}$$

Furthermore, also for $j = 1, 2$,

$$\sup_{\theta \in K} \left| \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_{\theta_3} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) - B_{j3}(\theta; \theta_0) \right| \xrightarrow{\mathcal{P}} 0 \,,$$

so, as the continuous function $\theta \mapsto B_{j3}(\theta; \theta_0)$ attains a maximum on K, also

$$\begin{aligned} &\sup_{\theta \in K} \left| \frac{1}{n\Delta_n^{1/2}} \sum_{i=1}^{n} \partial_{\theta_3} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \right| \\ &\quad \leq \Delta_n^{1/2} \left(\sup_{\theta \in K} \left| \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_{\theta_3} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) - B_{j3}(\theta; \theta_0) \right| + \sup_{\theta \in K} \left| B_{j3}(\theta; \theta_0) \right| \right) \\ &\quad \xrightarrow{\mathcal{P}} 0 \,. \end{aligned} \tag{3.5.7}$$

Let

$$B_0(\theta; \theta_0) = \begin{pmatrix} B_{11}(\theta; \theta_0) & B_{12}(\theta; \theta_0) & 0 \\ B_{21}(\theta; \theta_0) & B_{22}(\theta; \theta_0) & 0 \\ 0 & 0 & B_{33}(\theta; \theta_0) \end{pmatrix}.$$

Together, (3.5.6), (3.5.7) and Lemma 3.5.3.(i) imply that

$$\sup_{\theta \in K} \|\delta_n \partial_\theta G_n(\theta) \delta_n^{-1} - B_0(\theta; \theta_0)\| \xrightarrow{\mathcal{P}} 0 \,,$$

where, in particular,

$$B_0(\theta_0; \theta_0) = \begin{pmatrix} B_1(\theta_0; \theta_0) & 0 \\ 0 & B_2(\theta_0; \theta_0) \end{pmatrix}.$$

Now, (3.4.9) follows from Theorem 3.B.4.

Finally, by Lemmas 3.5.1.(iii) and 3.5.3.(ii),

$$\begin{aligned} \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_\alpha g_\alpha(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) &\xrightarrow{\mathcal{P}} B_1(\theta_0; \theta_0) \\ \frac{1}{n\Delta_n} \sum_{i=1}^{n} \partial_\beta g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) &\xrightarrow{\mathcal{P}} B_2(\theta_0; \theta_0) \\ \frac{1}{n\Delta_n} \sum_{i=1}^{n} g_\alpha g_\alpha^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) &\xrightarrow{\mathcal{P}} E_1(\theta_0; \theta_0) \\ \frac{1}{n\Delta_n^2} \sum_{i=1}^{n} g_\beta^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n) &\xrightarrow{\mathcal{P}} E_2(\theta_0; \theta_0) \,, \end{aligned}$$

as under the present conditions, $E_1(\theta_0; \theta_0)$ is equal to $(C(\theta_0; \theta_0)_{jk})_{j,k=1,2}$ of Lemma 3.5.1. An application of the continuous mapping theorem completes the proof. □

3.5.3 Proofs of Main Lemmas

This section contains the proofs of Lemmas 3.5.1, 3.5.2 and 3.5.3. In these proofs, the notation

$$g_j^{n,i} = g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \quad \text{and} \quad \mathbb{E}_{\theta_0}^{i-1}(\,\cdot\,) = \mathbb{E}_{\theta_0}\left(\,\cdot \mid X_{t_{i-1}^n}\right)$$

is sometimes used. Also, a martingale difference central limit theorem is utilised several times (Hall and Heyde, 1980, Corollary 3.1). For convenience, a version of the applicable result of Hall and Heyde, tailored specifically to the current setup, is stated in Section 3.B.1.

Proof of Lemma 3.5.1. The identity $A(\theta_0; \theta_0) = 0$ is clearly satisfied. In order to prove the rest of Lemma 3.5.1.(i), observe the following. For any $f(x; \theta) \in C_{0,0}^{\text{pol}}(\mathcal{X} \times \Theta)$, $\lambda_0 \in \Theta$ and compact, convex $K \subseteq \Theta$ with $\lambda_0 \in \operatorname{int} K$, there exist constants $C_K > 0$ such that

$$|f(x; \theta)| \leq C_K(1 + |x|^{C_K}) \tag{3.5.8}$$

for all $\theta \in \operatorname{int} K$ and $x \in \mathcal{X}$. Under Assumption 3.2.5.(v), which ensures finite moments of π_{θ_0}, the right-hand side of (3.5.8) can be used as a π_{θ_0}-integrable majorant of f. By the help of Lemma 3.A.1,

$$\partial_\theta \mathcal{L}_{\theta_0}(g(0, \theta))(x, x) = \mathcal{L}_{\theta_0}(\partial_\theta g(0, \theta))(x, x)\,.$$

Finally, by Lemma 3.A.8, for each $j, k = 1, \ldots, d$, the integrands in $A_j(\theta; \theta_0)$ and $C_{jk}(\theta; \theta_0)$ are $C_{1,2}^{\text{pol}}(\mathcal{X} \times \Theta)$-functions, while the integrand in $B_{jk}(\theta; \theta_0)$, which is the partial derivative with respect to θ_k of the integrand in $A_j(\theta; \theta_0)$, is a $C_{1,1}^{\text{pol}}(\mathcal{X} \times \Theta)$-function. Keeping in mind these considerations, the remaining results in Lemma 3.5.1.(i) follow by the usual results for continuity and differentiability of functions given by integrals (the dominated convergence theorem).

Now, in order to prove Lemma 3.5.1.(ii) and (iii), combine Lemmas 3.A.8 and 3.A.25 with Lemma 3.A.29 and Remark 3.A.30, to see that for $j, j_1, j_2, k = 1, \ldots, d$,

$$\begin{aligned}
&\frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{\theta_0}\left(g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \frac{1}{n} \sum_{i=1}^n \left(\mathcal{L}_{\theta_0}(g_j(0; \theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - \mathcal{L}_\theta(g_j(0; \theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right) \\
&\qquad + \Delta_n \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\quad \xrightarrow{\mathcal{P}} A_j(\theta; \theta_0)\,,
\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{n\Delta_n}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(\partial_{\theta_k}g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{L}_{\theta_0}(\partial_{\theta_k}g_j(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})-\partial_{\theta_k}\mathcal{L}_{\theta}(g_j(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)\\
&\qquad+\Delta_n\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad\xrightarrow{\mathcal{P}} B_{jk}(\theta;\theta_0)\,,\\
&\frac{1}{n\Delta_n}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)\\
&\quad=\frac{1}{n}\sum_{i=1}^{n}\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}(0,\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n})+\Delta_n\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad\xrightarrow{\mathcal{P}} C_{j_1j_2}(\theta;\theta_0)\,,
\end{aligned}
\tag{3.5.9}
$$

and

$$
\begin{aligned}
\frac{1}{(n\Delta_n)^2}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g_j^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)&=\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n};\theta)\xrightarrow{\mathcal{P}}0\,,\\
\frac{1}{(n\Delta_n)^2}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left((\partial_{\theta_k}g_j)^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)&=\frac{1}{n\Delta}\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n};\theta)\xrightarrow{\mathcal{P}}0\,,\\
\frac{1}{(n\Delta_n)^2}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g_{j_1}^2g_{j_2}^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right)&=\frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n};\theta)\xrightarrow{\mathcal{P}}0\,,
\end{aligned}
$$

implying that

$$
\begin{aligned}
A_j^{(n)}(\theta)&=\frac{1}{n\Delta_n}\sum_{i=1}^{n}g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\xrightarrow{\mathcal{P}}A_j(\theta;\theta_0)\\
B_{jk}^{(n)}(\theta)&=\frac{1}{n\Delta_n}\sum_{i=1}^{n}\partial_{\theta_k}g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\xrightarrow{\mathcal{P}}B_{jk}(\theta;\theta_0)\\
C_{j_1j_2}^{(n)}(\theta)&=\frac{1}{n\Delta_n}\sum_{i=1}^{n}g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\xrightarrow{\mathcal{P}}C_{j_1j_2}(\theta;\theta_0)
\end{aligned}
$$

point-wise for $\theta\in\Theta$ by Lemma 3.A.31.

Let any compact, convex set $K\subseteq\Theta$ be given. The functions $g_j(t,y,x;\theta)$, $\partial_{\theta_k}g_j(t,y,x;\theta)$ and $g_{j_1}g_{j_2}(t,y,x;\theta)$ all satisfy the conditions of Lemma 3.A.16, which may be used together with Lemma 3.A.22 to conclude the existence of constants $p>d$ and $C_{K,p}>0$ such that

$$
\begin{aligned}
\mathbb{E}_{\theta_0}\left(\left|A_j^{(n)}(\theta)-A_j(\theta;\theta_0)-A_j^{(n)}(\theta')+A_j(\theta';\theta_0)\right|^p\right)&\leq C_{K,p}\|\theta-\theta'\|^p\\
\mathbb{E}_{\theta_0}\left(\left|B_{jk}^{(n)}(\theta)-B_{jk}(\theta;\theta_0)-B_{jk}^{(n)}(\theta')+B_{jk}(\theta';\theta_0)\right|^p\right)&\leq C_{K,p}\|\theta-\theta'\|^p
\end{aligned}
$$

$$\mathbb{E}_{\theta_0}\left(\left|C^{(n)}_{j_1j_2}(\theta) - C_{j_1j_2}(\theta;\theta_0) - C^{(n)}_{j_1j_2}(\theta') + C_{j_1j_2}(\theta';\theta_0)\right|^p\right) \leq C_{K,p}\|\theta-\theta'\|^p .$$

Now, by Lemma 3.A.32, for all compact, convex sets $K \subseteq \Theta$,

$$\sup_{\theta\in K}\left|A^{(n)}_j(\theta) - A_j(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0$$
$$\sup_{\theta\in K}\left|B^{(n)}_{jk}(\theta) - B_{jk}(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0$$
$$\sup_{\theta\in K}\left|C^{(n)}_{j_1j_2}(\theta) - C_{j_1j_2}(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0 ,$$

and Lemma 3.5.1.(ii) follows. Lemma 3.5.1.(iii) is immediate by application of Lemma 3.A.33.

Finally, Lemma 3.5.1.(iv) is shown using Theorem 3.B.1 (Hall and Heyde, 1980, Corollary 3.1) and the Cramér-Wold device. Note first that by Lemma 3.A.25, used together with Remark 3.A.30, it holds that for $j_1, j_2, j_3, j_4 = 1, \ldots, d$,

$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n \mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_1}g^{n,i}_{j_2}g^{n,i}_{j_3}g^{n,i}_{j_4}\right) = \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t^n_{i-1}};\theta_0) \xrightarrow{\mathcal{P}} 0 , \tag{3.5.10}$$

and similarly, using also Lemmas 3.A.8 and 3.A.29,

$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n \mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_1}\right)\mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_2}g^{n,i}_{j_3}g^{n,i}_{j_4}\right) \xrightarrow{\mathcal{P}} 0$$
$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n \mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_1}\right)\mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_2}\right)\mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_3}g^{n,i}_{j_4}\right) \xrightarrow{\mathcal{P}} 0 \tag{3.5.11}$$
$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n \mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_1}\right)\mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_2}\right)\mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_3}\right)\mathbb{E}^{i-1}_{\theta_0}\left(g^{n,i}_{j_4}\right) \xrightarrow{\mathcal{P}} 0 .$$

Initially, suppose that the estimating function is a martingale estimating function, i.e. that $R_\theta(t,x) \equiv 0$ in Assumption 3.2.6.(i). Let $v \in \mathbb{R}^d$ be a fixed vector and consider

$$M_{n,i} = \frac{1}{\sqrt{n\Delta_n}}\sum_{j=1}^i v^\star g(\Delta_n, X_{t^n_j}, X_{t^n_{j-1}};\theta_0)$$

which constitutes a real-valued, zero-mean, square-integrable martingale array with differences $D_{n,i} = (n\Delta_n)^{-1/2}v^\star g(\Delta_n, X_{t^n_i}, X_{t^n_{i-1}};\theta_0)$, together with the σ-algebras $\mathcal{G}_{n,i}$ generated by $(X_{t^n_0}, \ldots, X_{t^n_i})$.

It holds that

$$\frac{1}{n\Delta_n}\sum_{i=1}^n \mathbb{E}_{\theta_0}\left((v^\star g)^2(\Delta_n, X_{t^n_i}, X_{t^n_{i-1}};\theta_0) \mid X_{t^n_i}\right)$$
$$= v^\star\left(\frac{1}{n\Delta_n}\sum_{i=1}^n \mathbb{E}_{\theta_0}\left(gg^\star(\Delta_n, X_{t^n_i}, X_{t^n_{i-1}};\theta_0) \mid X_{t^n_i}\right)\right)v$$

$$\xrightarrow{\mathcal{P}} v^\star C(\theta_0;\theta_0)v\,,$$

by (3.5.9), where $C(\theta_0;\theta_0)$ is a non-random matrix. Furthermore, the conditional Lyapunov condition

$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left((v^\star g)^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\mid X_{t_i^n}\right)\xrightarrow{\mathcal{P}} 0 \tag{3.5.12}$$

holds, implying the Lindeberg condition of Theorem 3.B.1. The convergence in (3.5.12) is seen because the left-hand side may be written as a sum of terms of the form (3.5.10) (multiplied by deterministic constants $v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4}$) for $j_1, j_2, j_3, j_4 = 1,\ldots,d$. It follows then, from Theorem 3.B.1, that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n} v^\star g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0, v^\star C(\theta_0;\theta_0)v\right).$$

Now, by the Cramér-Wold device, Lemma 3.5.1.(iv) follows for martingale estimating functions.

If the estimating function is not a martingale estimating function, i.e. if Assumption 3.2.6.(i) holds with $R_\theta(\Delta_n, X_{t_{i-1}^n})\not\equiv 0$, then $n\Delta_n^{2\kappa-1}\to 0$ as $n\to\infty$ for some $\kappa\geq 2$. Let

$$\tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) = g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) - \mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\mid X_{t_{i-1}^n}\right).$$

Since,

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\mid X_{t_{i-1}^n}\right) = \sqrt{n}\Delta_n^{\kappa-1/2}\frac{1}{n}\sum_{i=1}^{n}R_{\theta_0}(\Delta_n, X_{t_{i-1}^n})\xrightarrow{\mathcal{P}} 0$$

by Assumption 3.2.6.(i) and Remark 3.A.30, it remains to show that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}\tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\xrightarrow{\mathcal{D}}\mathcal{N}_d(0, C(\theta_0;\theta_0))\,. \tag{3.5.13}$$

Again, let $v\in\mathbb{R}^d$ be a fixed vector and consider

$$M_{n,i} = \frac{1}{\sqrt{n\Delta_n}}\sum_{j=1}^{i} v^\star\tilde{g}(\Delta_n, X_{t_j^n}, X_{t_{j-1}^n};\theta_0)$$

with martingale differences $D_{n,i} = (n\Delta_n)^{-1/2}v^\star\tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)$. Using Lemma 3.A.25 and (3.5.9),

$$\begin{aligned}
&\frac{1}{n\Delta_n}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left((v^\star\tilde{g})^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\mid X_{t_{i-1}^n}\right)\\
&\quad = v^\star\left(\frac{1}{n\Delta_n}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(\tilde{g}\tilde{g}^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\mid X_{t_{i-1}^n}\right)\right)v
\end{aligned}$$

$$
\begin{aligned}
&= v^\star \left(\frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{\theta_0} \left(g g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \right) v \\
&\quad - v^\star \left(\frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right)^\star \right) v \\
&= v^\star \left(\frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_{\theta_0} \left(g g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) - \Delta_n^3 \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}; \theta_0) \right) v \\
&\xrightarrow{\mathcal{P}} v^\star C(\theta_0; \theta_0) v \,.
\end{aligned}
$$

Furthermore, the conditional Lyapunov condition

$$
\frac{1}{(n\Delta_n)^2} \sum_{i=1}^n \mathbb{E}_{\theta_0} \left((v^\star \tilde{g})^4 (\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_i^n} \right) \xrightarrow{\mathcal{P}} 0 \tag{3.5.14}
$$

holds, because the left-hand side of (3.5.14) may be written as sums of terms of the forms (3.5.10) and (3.5.11) (multiplied by deterministic constants $v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4}$) for $j_1, j_2, j_3, j_4 = 1, \ldots, d$. Now, by Theorem 3.B.1 and the Cramér-Wold device, (3.5.13) follows, thus Lemma 3.5.1.(iv) is also proved for approximate (non-exact) martingale estimating functions. □

Proof of Lemma 3.5.2. First, use Lemmas 3.A.26.(ii), 3.A.27.(iii) and 3.A.8 together with Lemma 3.A.29 and Remark 3.A.30 to see that

$$
\begin{aligned}
&\frac{1}{n\Delta_n^2} \sum_{i=1}^n \mathbb{E}_{\beta_0} \left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta) \mid X_{t_{i-1}^n} \right) \\
&= \frac{1}{n} \sum_{i=1}^n \tfrac{1}{2} \left(b^4(X_{t_{i-1}^n}; \beta_0) + \tfrac{1}{2} \left(b^2(X_{t_{i-1}^n}; \beta_0) - b^2(X_{t_{i-1}^n}; \beta) \right)^2 \right) \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \beta)^2 \\
&\quad + \Delta_n \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}; \beta) \\
&\xrightarrow{\mathcal{P}} D(\beta; \beta_0)
\end{aligned} \tag{3.5.15}
$$

and

$$
\frac{1}{n^2\Delta_n^4} \sum_{i=1}^n \mathbb{E}_{\beta_0} \left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta) \mid X_{t_{i-1}^n} \right) = \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}; \beta) \xrightarrow{\mathcal{P}} 0 \,, \tag{3.5.16}
$$

yielding

$$
\frac{1}{n\Delta_n^2} \sum_{i=1}^n g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta) - D(\beta; \beta_0) \xrightarrow{\mathcal{P}} 0 \,, \tag{3.5.17}
$$

pointwise for $\beta \in B$, by Lemma 3.A.31.

In order to prove Lemma 3.5.2.(i), note that by the arguments similar to those in the proof of Lemma 3.5.1.(i), $\beta \mapsto D(\beta; \beta_0)$ is continuous on B. Also, $g^2(t, y, x; \beta)$ satisfies the conditions on f in Lemma 3.A.20, and combining this with Lemma 3.A.22 and (3.5.17), it follows from Lemma 3.A.32 that

$$\sup_{\beta \in K} \left| \frac{1}{n\Delta_n^2} \sum_{i=1}^{n} g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta) - D(\beta; \beta_0) \right| \xrightarrow{\mathcal{P}} 0$$

for any compact, convex set $K \subseteq B$. Now, the convergence in Lemma 3.5.2.(i) follows from Lemma 3.A.33.

Lemma 3.5.2.(ii) is shown using Theorem 3.B.1 (Hall and Heyde, 1980, Corollary 3.1). Note, for use in the following, that by (3.A.78) and Lemmas 3.A.26 and 3.A.27,

$$\begin{aligned} \frac{1}{n^2\Delta_n^4} \sum_{i=1}^{n} \mathbb{E}_{\beta_0}^{i-1}\left(g^{n,i} g^{n,i} g^{n,i}\right) \mathbb{E}_{\beta_0}^{i-1}\left(g^{n,i}\right) \xrightarrow{\mathcal{P}} 0 \\ \frac{1}{n^2\Delta_n^4} \sum_{i=1}^{n} \mathbb{E}_{\beta_0}^{i-1}\left(g^{n,i} g^{n,i}\right) \mathbb{E}_{\beta_0}^{i-1}\left(g^{n,i}\right)^2 \xrightarrow{\mathcal{P}} 0 \\ \frac{1}{n^2\Delta_n^4} \sum_{i=1}^{n} \mathbb{E}_{\beta_0}^{i-1}\left(g^{n,i}\right)^4 \xrightarrow{\mathcal{P}} 0\,. \end{aligned} \tag{3.5.18}$$

Suppose that the estimating function is a martingale estimating function, i.e. $R_\beta(t,x) \equiv 0$ in Assumption 3.2.6.(i). Consider, for $n \in \mathbb{N}$,

$$M_{n,i} = \frac{1}{\sqrt{n}\Delta_n} \sum_{j=1}^{i} g(\Delta_n, X_{t_j^n}, X_{t_{j-1}^n}; \beta_0)\,,$$

which constitutes a real-valued, zero-mean, square-integrable martingale array with differences $D_{n,i} = n^{-1/2}\Delta_n^{-1} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0)$. Firstly, by (3.5.15), it holds that

$$\frac{1}{n\Delta_n^2} \sum_{i=1}^{n} \mathbb{E}_{\beta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right) \xrightarrow{\mathcal{P}} D(\beta_0; \beta_0)\,.$$

Secondly, by (3.5.16), the conditional Lyapunov condition

$$\frac{1}{n^2\Delta_n^4} \sum_{i=1}^{n} \mathbb{E}_{\beta_0}\left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right) \xrightarrow{\mathcal{P}} 0$$

holds, implying the Lindeberg condition of Theorem 3.B.1, so (3.5.2) follows in the case of a martingale estimating function. When the estimating function is not an (exact) martingale estimating function, i.e. Assumption 3.2.6.(i) holds with $R_\beta(t,x) \not\equiv 0$, and $n\Delta_n^{2(\kappa-1)} \to 0$ as $n \to \infty$ for some $\kappa \geq 2$, let

$$\tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) = g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) - \mathbb{E}_{\beta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right)\,.$$

As

$$\frac{1}{\sqrt{n}\Delta_n} \sum_{i=1}^{n} \mathbb{E}_{\beta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right) = \sqrt{n}\Delta_n^{\kappa-1} \frac{1}{n} \sum_{i=1}^{n} R_{\beta_0}(\Delta_n, X_{t_{i-1}^n}) \xrightarrow{\mathcal{P}} 0$$

by Remark 3.A.30, it remains to show that

$$\frac{1}{\sqrt{n}\Delta_n}\sum_{i=1}^{n}\tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, D(\beta_0;\beta_0))\,. \tag{3.5.19}$$

First, see that

$$\begin{aligned}
&\frac{1}{n\Delta_n^2}\sum_{i=1}^{n}\mathbb{E}_{\beta_0}\left(\tilde{g}^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right)\\
&\quad= \frac{1}{n\Delta_n^2}\sum_{i=1}^{n}\left(\mathbb{E}_{\beta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right) - \mathbb{E}_{\beta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right)^2\right)\\
&\quad= \frac{1}{n\Delta_n^2}\sum_{i=1}^{n}\mathbb{E}_{\beta_0}\left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right) - \Delta_n^2\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n}; \beta_0)\\
&\quad\xrightarrow{\mathcal{P}} D(\beta_0;\beta_0)\,.
\end{aligned}$$

Then, observe that the conditional Lyapunov condition

$$\frac{1}{n^2\Delta_n^4}\sum_{i=1}^{n}\mathbb{E}_{\beta_0}\left(\tilde{g}^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \beta_0) \mid X_{t_{i-1}^n}\right) \xrightarrow{\mathcal{P}} 0 \tag{3.5.20}$$

holds, since the left-hand side of (3.5.20) may be written as a sum of terms of the form (3.5.16) and (3.5.18). This implies the Lindeberg condition of Theorem 3.B.1. It follows then, that (3.5.19) holds, thus completing the proof of Lemma 3.5.2.(ii). □

Proof of Lemma 3.5.3. Let

$$E_2(\theta;\theta_0) = \int_{\mathcal{X}}\tfrac{1}{2}\left(b^4(x;\beta_0) + \tfrac{1}{2}\left(b^2(x;\beta_0) - b^2(x;\beta)\right)^2\right)\partial_y^2 g_3(0, x, x, \theta)^2\,\pi_{\theta_0}(dx)\,.$$

First, use Lemma 3.A.29 and Remark 3.A.30 together with Lemmas 3.A.26, 3.A.27 and 3.A.8 to see that

$$\begin{aligned}
&\frac{1}{n\Delta_n^2}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g_3^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right)\\
&= \frac{1}{n}\sum_{i=1}^{n}\tfrac{1}{2}\left(b^4(X_{t_{i-1}^n};\beta_0) + \tfrac{1}{2}\left(b^2(X_{t_{i-1}^n};\beta_0) - b^2(X_{t_{i-1}^n};\beta)\right)^2\right)\partial_y^2 g_3(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta)^2\\
&\quad + \Delta_n\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n}; \theta)\\
&\xrightarrow{\mathcal{P}} E_2(\theta;\theta_0)
\end{aligned} \tag{3.5.21}$$

and

$$\frac{1}{n^2\Delta_n^4}\sum_{i=1}^{n}\mathbb{E}_{\theta_0}\left(g_3^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) = \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n, X_{t_{i-1}^n}; \theta) \xrightarrow{\mathcal{P}} 0$$

$$\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n \mathbb{E}_{\theta_0}\left(g_j g_3(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^{1/2}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_i^n};\theta) \xrightarrow{\mathcal{P}} 0 \tag{3.5.22}$$

for $j = 1,2$, and together with Lemma 3.A.28 to see that

$$\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n \mathbb{E}_{\theta_0}\left(\partial_{\alpha_j} g_3(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^{1/2}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta) \xrightarrow{\mathcal{P}} 0\,,$$
$$\frac{1}{n^2\Delta_n^{3}}\sum_{i=1}^n \mathbb{E}_{\theta_0}\left(\partial_{\alpha_j} g_3(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right) = \frac{1}{n}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta) \xrightarrow{\mathcal{P}} 0\,.$$

Then, for $j = 1,2$,

$$\frac{1}{n\Delta_n^{2}}\sum_{i=1}^n g_3^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) - E_2(\theta;\theta_0) \xrightarrow{\mathcal{P}} 0 \tag{3.5.23}$$
$$\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n \partial_{\alpha_j} g_3(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \xrightarrow{\mathcal{P}} 0 \tag{3.5.24}$$

pointwise for $\theta \in \Theta$ by Lemma 3.A.31. The function $\partial_{\alpha_j} g_3(t,y,x;\theta)$ satisfies the conditions on f in Lemma 3.A.19, so Lemma 3.5.3.(i) follows by (3.5.24) and Lemma 3.A.32. Furthermore, $g_3^2(t,y,x;\theta)$ satisfies the conditions on f in Lemma 3.A.21, so

$$\sup_{\theta\in K}\left|\frac{1}{n\Delta_n^{2}}\sum_{i=1}^n g_3^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) - E_2(\theta;\theta_0)\right| \xrightarrow{\mathcal{P}} 0$$

by (3.5.23) and Lemmas 3.A.22 and 3.A.32. Now, Lemma 3.5.3.(ii) follows from Lemma 3.A.33.

In order to prove Lemma 3.5.3.(iii), observe first that

$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n \delta_n \mathbb{E}_{\theta_0}\left(g g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right)\delta_n$$
$$= \begin{pmatrix} \frac{1}{n\Delta_n}\sum_{i=1}^n \mathbb{E}_{\theta_0}^{i-1}\left(g_\alpha g_\alpha^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\right) & \frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n \mathbb{E}_{\theta_0}^{i-1}\left(g_\alpha g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\right) \\ \frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n \mathbb{E}_{\theta_0}^{i-1}\left(g_\beta g_\alpha^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\right) & \frac{1}{n\Delta_n^{2}}\sum_{i=1}^n \mathbb{E}_{\theta_0}^{i-1}\left(g_\beta^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\right) \end{pmatrix},$$

so combining (3.5.9) and Remark 3.2.10 for the submatrix concerning $g_\alpha g_\alpha^\star$, and (3.5.21) and (3.5.22) for the remaining coordinates, it follows that

$$\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n \delta_n \mathbb{E}_{\theta_0}\left(g g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \mid X_{t_{i-1}^n}\right)\delta_n \xrightarrow{\mathcal{P}} E(\theta_0;\theta_0)\,. \tag{3.5.25}$$

Also, for $j_1, j_2, j_3, j_4 = 1, 2$,

$$\begin{aligned}
\frac{1}{(n\Delta_n)^2} \sum_{i=1}^{n} \mathbb{E}_{\theta_0}^{i-1}\left(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{j_3}^{n,i} g_{j_4}^{n,i}\right) &\xrightarrow{\mathcal{P}} 0 \\
\frac{1}{n^2\Delta_n^{5/2}} \sum_{i=1}^{n} \mathbb{E}_{\theta_0}^{i-1}\left(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{j_3}^{n,i} g_{3}^{n,i}\right) &\xrightarrow{\mathcal{P}} 0 \\
\frac{1}{n^2\Delta_n^{3}} \sum_{i=1}^{n} \mathbb{E}_{\theta_0}^{i-1}\left(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{3}^{n,i} g_{3}^{n,i}\right) &\xrightarrow{\mathcal{P}} 0 \qquad (3.5.26)\\
\frac{1}{n^2\Delta_n^{7/2}} \sum_{i=1}^{n} \mathbb{E}_{\theta_0}^{i-1}\left(g_{j_1}^{n,i} g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i}\right) &\xrightarrow{\mathcal{P}} 0 \\
\frac{1}{n^2\Delta_n^{4}} \sum_{i=1}^{n} \mathbb{E}_{\theta_0}^{i-1}\left(g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i}\right) &\xrightarrow{\mathcal{P}} 0
\end{aligned}$$

by (3.A.79), Lemma 3.A.27 and Remark 3.A.30.

Suppose now that we're dealing with a martingale estimating function, i.e. $R_\theta(t, x) \equiv 0$ in Assumption 3.2.6.(i). Let $v \in \mathbb{R}^3$ be a fixed vector, and

$$M_{n,i} = \frac{1}{n\Delta_n} \sum_{j=1}^{i} v^\star \delta_n g(\Delta_n, X_{t_j^n}, X_{t_{j-1}^n}; \theta_0)$$

be the variables in a real-valued, zero-mean, square-integrable $\mathcal{G}_{n,i}$-martingale array with differences $D_{n,i} = (n\Delta_n)^{-1} v^\star \delta_n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)$. By (3.5.25), it holds that

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left(\left((n\Delta_n)^{-1} v^\star \delta_n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)\right)^2 \mid X_{t_{i-1}^n}\right) \\
&\quad = v^\star \left(\frac{1}{(n\Delta_n)^2} \sum_{i=1}^{n} \delta_n \mathbb{E}_{\theta_0}\left(g g^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n}\right) \delta_n\right) v \\
&\quad \xrightarrow{\mathcal{D}} v^\star E(\theta_0; \theta_0) v\,.
\end{aligned}$$

Furthermore, the conditional Lyapunov condition

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left(\left((n\Delta_n)^{-1} v^\star \delta_n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)\right)^4 \mid X_{t_i^n}\right) \\
&\quad = \sum_{i=1}^{n} \mathbb{E}_{\theta_0}^{i-1}\left(\left(\tfrac{1}{\sqrt{n\Delta_n}}\left(v_1 g_1^{n,i} + v_2 g_2^{n,i}\right) + \tfrac{1}{\sqrt{n}\Delta_n} v_3 g_3^{n,i}\right)^4\right) \qquad (3.5.27)\\
&\quad \xrightarrow{\mathcal{P}} 0
\end{aligned}$$

holds, implying the Lindeberg condition of Theorem 3.B.1 (Hall and Heyde, 1980, Corollary 3.1). The convergence in (3.5.27) holds, because the second sum may be written as a sum of terms of the form (3.5.26) for $j_1, j_2, j_3, j_4 = 1, 2$ (omitting constant factors v_1, v_2 and v_3). It follows then, that

$$\frac{1}{n\Delta_n} \sum_{i=1}^{n} v^\star \delta_n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, v^\star E(\theta_0; \theta_0) v\right)\,,$$

thus proving Lemma 3.5.3.(iii) for martingale estimating functions, by the Cramér-Wold device.

If we're not dealing with a martingale estimating function, in which case $n\Delta_n^{2(\kappa-1)} \to 0$ as $n \to \infty$ for some $\kappa \geq 2$, let

$$\tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) = g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) - \mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n}\right).$$

Since

$$\delta_n \Delta_n^{\kappa-1} = \begin{pmatrix} n^{1/2}\Delta_n^{\kappa-1/2} & 0 & 0 \\ 0 & n^{1/2}\Delta_n^{\kappa-1/2} & 0 \\ 0 & 0 & n^{1/2}\Delta_n^{\kappa-1} \end{pmatrix} \to 0$$

as $n \to \infty$, it holds that

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n} \delta_n \mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n}\right) = \delta_n \Delta_n^{\kappa-1} \frac{1}{n}\sum_{i=1}^{n} R(\Delta_n, X_{t_{i-1}^n}; \theta_0) \xrightarrow{\mathcal{P}} 0,$$

and it remains to show that

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n} \delta_n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_3(0, E(\theta_0; \theta_0)). \tag{3.5.28}$$

Again, let $v \in \mathbb{R}^3$ be a fixed vector and consider

$$M_{n,i} = \frac{1}{n\Delta_n}\sum_{j=1}^{i} v^\star \delta_n \tilde{g}(\Delta_n, X_{t_j^n}, X_{t_{j-1}^n}; \theta_0)$$

with martingale differences $D_{n,i} = (n\Delta_n)^{-1} v^\star \delta_n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)$. Then,

$$\begin{aligned}
&\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left(\left((n\Delta_n)^{-1} v^\star \delta_n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)\right)^2 \mid X_{t_{i-1}^n}\right) \\
&\quad = v^\star \left(\frac{1}{(n\Delta_n)^2}\sum_{i=1}^{n} \delta_n \mathbb{E}_{\theta_0}\left(gg^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n}\right)\delta_n\right) v \\
&\qquad - v^\star \delta_n \Delta_n^{\kappa-1}\left(\frac{1}{n^2}\sum_{i=1}^{n} R(\Delta_n, X_{t_{i-1}^n}; \theta_0)\right)\delta_n \Delta_n^{\kappa-1} v \\
&\quad \xrightarrow{\mathcal{P}} v^\star E(\theta_0; \theta_0) v.
\end{aligned}$$

Also, the conditional Lyapunov condition

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left(\left((n\Delta_n)^{-1} v^\star \delta_n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)\right)^4 \mid X_{t_i^n}\right) \xrightarrow{\mathcal{P}} 0 \tag{3.5.29}$$

holds. In order to see this, write

$$(n\Delta_n)^{-1} v^\star \delta_n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)$$

$$
\begin{aligned}
&= \tfrac{1}{\sqrt{n\Delta_n}}\left(v_1 g_1^{n,i} + v_2 g_2^{n,i}\right) + \tfrac{1}{\sqrt{n\Delta_n}} v_3 g_3^{n,i} \\
&\quad - \tfrac{1}{\sqrt{n\Delta_n}}\left(v_1 \mathbb{E}_{\theta_0}^{i-1}\left(g_1^{n,i}\right) + v_2 \mathbb{E}_{\theta_0}^{i-1}\left(g_2^{n}\right)\right) - \tfrac{1}{\sqrt{n\Delta_n}} v_3 \mathbb{E}_{\theta_0}^{i-1}\left(g_3^{n,i}\right)
\end{aligned}
$$

and define, for $j_1, j_2, j_3, j_4 = 1, 2$, terms of *type* $T_{i,n}^{(0)}$ to be of the form

$$
\begin{array}{ll}
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{j_3}^{n,i} g_{j_4}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_4}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{j_3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_4}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_4}^{n,i}) \,,
\end{array}
$$

terms of *type* $T_{i,n}^{(1)}$ to be of the form

$$
\begin{array}{ll}
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{j_3}^{n,i} g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{j_3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \,,
\end{array}
$$

terms of *type* $T_{i,n}^{(2)}$ to be of the form

$$
\begin{array}{ll}
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{3}^{n,i} g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{3}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i} g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{j_2}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \,, &
\end{array}
$$

terms of *type* $T_{i,n}^{(3)}$ to be of the form

$$
\begin{array}{ll}
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{3}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{j_1}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \,,
\end{array}
$$

and, finally, terms of *type* $T_{i,n}^{(4)}$ to be of the form

$$
\begin{array}{ll}
\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i} g_{3}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \\
\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i} g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) & \mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i})\mathbb{E}_{\theta_0}^{i-1}(g_{3}^{n,i}) \,.
\end{array}
$$

Using expressions for conditional moments from Lemmas 3.A.25, 3.A.26 and 3.A.27, together with Lemmas 3.A.8 and 3.A.29, and Remark 3.A.30, it may be verified that

$$
\frac{1}{n^2 \Delta_n^{2+k/2}} \sum_{i=1}^{n} T_{i,n}^{(k)} \xrightarrow{\mathcal{P}} 0 \quad \text{for} \quad k = 0, 1, 2, 3, 4 \,. \tag{3.5.30}
$$

The left-hand side of (3.5.29) may be written as sums of terms of the form (3.5.30) (multiplied by deterministic constants $v_{j_1}, v_{j_2}, v_{j_3}, v_{j_4}$). Now, by Theorem 3.B.1 and the Cramér-Wold device, (3.5.28) follows, thus completing the proof of Lemma 3.5.3.(iii). □

Appendix

3.A Auxiliary Results

This appendix contains technical results, mostly pertaining to the proofs of the main lemmas given in Section 3.5.3. The lemmas in Appendix 3.A.1 essentially verify that integrals with respect to the Lévy measure inherit polynomial growth properties of the integrand. Appendix 3.A.2 gives expressions for the infinitesimal generator applied to various functions. Appendix 3.A.3 contains inequalities to do with expectations, most of them used to prove uniform convergence in probability. Appendix 3.A.4 contains a number of expansions of conditional moments, as well as the proof of the expansion lemma, Lemma 3.2.8. Finally, Appendix 3.A.5 states some results on convergence in probability.

3.A.1 Polynomial Growth

Lemma 3.A.1. *Suppose that Assumption 3.2.5 holds, that*

$$f(t,y,x;\theta) \in C^{pol}_{p,q,r,s}((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$$

for some $p,q,r,s \in \mathbb{N}_0$, *and that* $c(y,z;\theta) \in C^{p\text{-}pol}_{q,0}(\mathcal{X} \times \mathbb{R} \times \Theta)$. *Let* $\lambda \in \Theta$ *be given, and define*

$$\phi_\lambda(t,y,x;\theta) = \int_{\mathbb{R}} f(t, y + c(y,z;\lambda), x;\theta)\, \nu_\lambda(dz)\,.$$

Then $\phi_\lambda(t,y,x;\theta) \in C^{pol}_{p,q,r,s}((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ *with*

$$\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \phi_\lambda(t,y,x;\theta) = \int_{\mathbb{R}} \partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l (f(t, y + c(y,z;\lambda), x;\theta))\, \nu_\lambda(dz)$$

for $i = 0,\dots,p$, $j = 0,\dots,q$, $k = 0,\dots,r$, $l = 0,\dots,s$ *and* $m = 1,\dots,d$. ◇

Proof of Lemma 3.A.1. Let i, j, k, l and m be given in the following, and introduce the notation $\tilde{f}_\lambda(t,x,y,z;\theta) = f(t, y+c(y,z;\lambda), x;\theta)$ and $h_\lambda(y,z) = y + c(y,z;\lambda)$, so that $\tilde{f}_\lambda(t,x,y,z;\theta) = f(t, h_\lambda(y,z), x;\theta)$. By the chain rule for higher order derivatives (also known as *Faá di Bruno's formula*),

$$\begin{aligned} &\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda(t,x,y,z;\theta) \\ &= \sum_{(\eta_1,\dots,\eta_j)\in M_j} \frac{j!}{\eta_1!\cdots\eta_j!} \left(\frac{\partial_y h_\lambda(y,z)}{1!}\right)^{\eta_1} \cdots \left(\frac{\partial_y^j h_\lambda(y,z)}{j!}\right)^{\eta_j} \partial_t^i \partial_y^\eta \partial_x^k \partial_{\theta_m}^l f(t, h_\lambda(y,z), x;\theta) \end{aligned} \tag{3.A.1}$$

where $\eta = \eta_1 + \cdots + \eta_j$ and $M_j = \{x \in \mathbb{N}_0^j \mid x_1 + 2x_2 + \cdots + jx_j = j\}$.

Using the product-polynomial growth properties of $c(x, z; \lambda)$, it may be verified that for each $\eta = 0, \ldots, j$ and any compact, convex set $K \subseteq \Theta$, there exist constants $C, C_K > 0$ such that

$$\left|\partial_t^i \partial_y^\eta \partial_x^k \partial_{\theta_m}^l f(t, y + c(y, z; \lambda), x; \theta)\right| \leq C_K \left(1 + |x|^{C_K} + |y|^{C_K}\right)\left(1 + |z|^{C_K}\right)$$

and

$$\left|\partial_y^\eta c(y, z; \lambda)\right| \leq C\left(1 + |y|^C\right)\left(1 + |z|^C\right) \leq C\left(1 + |x|^C + |y|^C\right)\left(1 + |z|^C\right)$$

for all $t \in (0, \Delta_0)_{\varepsilon_0}$, $x, y \in \mathcal{X}$, $z \in \mathbb{R}$ and $\theta \in K$. Now, (3.A.1) may be used to conclude that there exist constants $C_K > 0$ such that

$$|\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda(t, x, y, z; \theta)| \leq C_K \left(1 + |x|^{C_K} + |y|^{C_K}\right)\left(1 + |z|^{C_K}\right), \tag{3.A.2}$$

hence

$$\begin{aligned}
&\int_{\mathbb{R}} |\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda(t, x, y, z; \theta)|\, \nu_\lambda(dz) \\
&\quad \leq C_K \left(1 + |x|^{C_K} + |y|^{C_K}\right) \int_{\mathbb{R}} \left(1 + |z|^{C_K}\right) \nu_\lambda(dz) \\
&\quad \leq C_K \left(1 + |x|^{C_K} + |y|^{C_K}\right)
\end{aligned}$$

for $t \in (0, \Delta_0)_{\varepsilon_0}$, $x, y \in \mathcal{X}$, $z \in \mathbb{R}$ and $\theta \in K$, showing that the function

$$(y, x; \theta) \mapsto \int_{\mathbb{R}} \partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda(t, x, y, z; \theta)\, \nu_\lambda(dz)$$

is well-defined and of polynomial growth in x and y, uniformly for t in $(0, \Delta_0)_{\varepsilon_0}$ and θ in compact, convex sets, for $x, y \in \mathcal{X}$. It is also seen by (3.A.1) that the partial derivatives $\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda$ are continuous in $(t, x, y, z; \theta)$.

Now, for any choice of $t_0 \in (0, \Delta_0)_{\varepsilon_0}$, $x_0, y_0 \in \mathcal{X}$ and $\lambda_0 \in \Theta$, choose $\varepsilon > 0$ such that $[x_0 - \varepsilon, x_0 + \varepsilon] \times [y_0 - \varepsilon, y_0 + \varepsilon] \subseteq \mathcal{X}^2$, and a compact, convex set $K \subseteq \Theta$ with $\lambda_0 \in \operatorname{int} K$. Recall from (3.A.2) that there exist constants $C_K^0 > 0$ such that

$$|\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda(t, x, y, z; \theta)| \leq C_K^0 \left(1 + |x|^{C_K^0} + |y|^{C_K^0}\right)\left(1 + |z|^{C_K^0}\right) \tag{3.A.3}$$

for $t \in (0, \Delta_0)_{\varepsilon_0}$, $x, y \in \mathcal{X}$, $z \in \mathbb{R}$ and $\theta \in K$. Let (x^*, y^*) denote the point where the factor $C_K^0(1 + |x|^{C_K^0} + |y|^{C_K^0})$ on the right-hand side of (3.A.3) achieves its maximum value on $[x_0 - \varepsilon, x_0 + \varepsilon] \times [y_0 - \varepsilon, y_0 + \varepsilon]$ as a function of (x, y). Now, using the same constants $C_K^0 > 0$ as in (3.A.3), the function $u : \mathbb{R} \to (0, \infty)$ defined by

$$u(z) = C_K^0 \left(1 + |x^*|^{C_K^0} + |y^*|^{C_K^0}\right)\left(1 + |z|^{C_K^0}\right)$$

is an integrable upper bound for $\partial_t^i \partial_y^j \partial_x^k \partial_{\theta_m}^l \tilde{f}_\lambda(t, x, y, z; \theta)$, for (t, x, y, θ) in the open set $(0, \Delta_0)_{\varepsilon_0} \times (x_0 - \varepsilon, x_0 + \varepsilon) \times (y_0 - \varepsilon, y_0 + \varepsilon) \times \operatorname{int} K$. This method of constructing integrable upper bounds for open neighbourhoods of any $(t_0, x_0, y_0, \lambda_0)$ may be used to conclude the desired continuity and differentiability results by the dominated convergence theorem. □

Lemma 3.A.2. *Suppose that Assumption 3.2.5 holds, and that*

$$f(y, x; \theta) \in C^{pol}_{3,1,2}(\mathcal{X}^2 \times \Theta) .$$

Define

$$\phi(y, x; \theta) = \int_{\mathbb{R}} f(y + c(y, z; \theta), x; \theta)\, \nu_\theta(dz) .$$

Then $\phi(y, x; \theta) \in C^{pol}_{1,1,2}(\mathcal{X}^2 \times \Theta)$ *with*

$$\partial_y^j\, \partial_x^k \partial_{\theta_m}^l\, \phi(y, x; \theta) = \int_{\mathbb{R}} \partial_y^j\, \partial_x^k\, \partial_{\theta_m}^l (f(y + c(y, z; \theta), x; \theta) q(z; \theta))\, \tilde{\nu}(dz)$$

for $j, k = 0, 1,\ l = 0, 1, 2$ *and* $m = 1, \dots, d$. ⋄

Lemma 3.A.2 involves, in a sense, more complicated derivatives than Lemma 3.A.1, as the fixed λ is replaced by the variable θ. Therefore, regarding the order of the derivatives, a less general result is stated in this case, tailored to fit the needs of this paper. The result is easily verified by differentiation, and the creation of upper bounds similar to those in the proof of Lemma 3.A.1. Assumption 3.2.5.(vii) is used to deal with the derivatives of the Lévy density.

3.A.2 The Infinitesimal Generator

In the following, expressions for the infinitesimal generator, sometimes applied twice, need to be computed several times for the products of two or more functions. For convenience, some general formulae are derived first.

For $f(y)$ and $h(y)$, functions of one variable, differentiable as often as necessary, the generalised Leibnitz formula gives:

$$\partial_y^m (fh)(y) = \sum_{k=0}^{m} \binom{m}{k} \partial_y^k f(y) \partial_y^{m-k} h(y) . \tag{3.A.4}$$

In particular, by (3.A.4), the first four derivatives of the product fh may be written as

$$\begin{aligned}
\partial_y(fh) &= f\partial_y h + \partial_y f h \\
\partial_y^2(fh) &= f\partial_y^2 h + 2\partial_y f \partial_y h + \partial_y^2 f h \\
\partial_y^3(fh) &= f\partial_y^3 h + 3\partial_y f \partial_y^2 h + 3\partial_y^2 f \partial_y h + \partial_y^3 f h \\
\partial_y^4(fh) &= f\partial_y^4 h + 4\partial_y f \partial_y^3 h + 6\partial_y^2 f \partial_y^2 h + 4\partial_y^3 f \partial_y h + \partial_y^4 f h .
\end{aligned} \tag{3.A.5}$$

Furthermore, if $f = f_1 f_2$, omitting combinatorial constants, the first two derivatives of the product fh may also be written as sums of terms of the following form

$$\begin{array}{llll}
\partial_y(f_1 f_2 h): & f_1 f_2 \partial_y h & f_1 \partial_y f_2 h & \partial_y f_1 f_2 h \\
\partial_y^2(f_1 f_2 h): & f_1 f_2 \partial_y^2 h & f_1 \partial_y f_2 \partial_y h & \partial_y f_1 f_2 \partial_y h \\
 & f_1 \partial_y^2 f_2 h & \partial_y f_1 \partial_y f_2 h & \partial_y^2 f_1 f_2 h .
\end{array} \tag{3.A.6}$$

and similarly, if also $h = h_1 h_2$,

$$
\begin{array}{llll}
\partial_y(f_1 f_2 h_1 h_2): & f_1 f_2 h_1 \partial_y h_2 & f_1 f_2 \partial_y h_1 h_2 & f_1 \partial_y f_2 h_1 h_2 \\
 & \partial_y f_1 f_2 h_1 h_2 & & \\
\partial_y^2(f_1 f_2 h_1 h_2): & f_1 f_2 h_1 \partial_y^2 h_2 & f_1 f_2 \partial_y h_1 \partial_y h_2 & f_1 f_2 \partial_y^2 h_1 h_2 \\
 & f_1 \partial_y f_2 h_1 \partial_y h_2 & f_1 \partial_y f_2 \partial_y h_1 h_2 & \partial_y f_1 f_2 h_1 \partial_y h_2 \\
 & \partial_y f_1 f_2 \partial_y h_1 h_2 & f_1 \partial_y^2 f_2 h_1 h_2 & \partial_y f_1 \partial_y f_2 h_1 h_2 \\
 & \partial_y^2 f_1 f_2 h_1 h_2 & & \\
\partial_y^3(f_1 f_2 h_1 h_2): & f_1 f_2 h_1 \partial_y^3 h_2 & f_1 f_2 \partial_y h_1 \partial_y^2 h_2 & f_1 f_2 \partial_y^2 h_1 \partial_y h_2 \\
 & f_1 f_2 \partial_y^3 h_1 h_2 & f_1 \partial_y f_2 h_1 \partial_y^2 h_2 & f_1 \partial_y f_2 \partial_y h_1 \partial_y h_2 \\
 & f_1 \partial_y f_2 \partial_y^2 h_1 h_2 & \partial_y f_1 f_2 h_1 \partial_y^2 h_2 & \partial_y f_1 f_2 \partial_y h_1 \partial_y h_2 \\
 & \partial_y f_1 f_2 \partial_y^2 h_1 h_2 & f_1 \partial_y^2 f_2 h_1 \partial_y h_2 & \partial_y f_1 \partial_y f_2 h_1 \partial_y h_2 \\
 & \partial_y^2 f_1 f_2 h_1 \partial_y h_2 & f_1 \partial_y^2 f_2 \partial_y h_1 h_2 & \partial_y f_1 \partial_y f_2 \partial_y h_1 h_2 \\
 & \partial_y^2 f_1 f_2 \partial_y h_1 h_2 & f_1 \partial_y^3 f_2 h_1 h_2 & \partial_y f_1 \partial_y^2 f_2 h_1 h_2 \\
 & \partial_y^2 f_1 \partial_y f_2 h_1 h_2 & \partial_y^3 f_1 f_2 h_1 h_2 & \\
\partial_y^4(f_1 f_2 h_1 h_2): & f_1 f_2 h_1 \partial_y^4 h_2 & f_1 f_2 \partial_y h_1 \partial_y^3 h_2 & f_1 f_2 \partial_y^2 h_1 \partial_y^2 h_2 \\
 & f_1 f_2 \partial_y^3 h_1 \partial_y h_2 & f_1 f_2 \partial_y^4 h_1 h_2 & f_1 \partial_y f_2 h_1 \partial_y^3 h_2 \\
 & f_1 \partial_y f_2 \partial_y h_1 \partial_y^2 h_2 & f_1 \partial_y f_2 \partial_y^2 h_1 \partial_y h_2 & f_1 \partial_y f_2 \partial_y^3 h_1 h_2 \\
 & \partial_y f_1 f_2 h_1 \partial_y^3 h_2 & \partial_y f_1 f_2 \partial_y h_1 \partial_y^2 h_2 & \partial_y f_1 f_2 \partial_y^2 h_1 \partial_y h_2 \\
 & \partial_y f_1 f_2 \partial_y^3 h_1 h_2 & f_1 \partial_y^2 f_2 h_1 \partial_y^2 h_2 & f_1 \partial_y^2 f_2 \partial_y h_1 \partial_y h_2 \\
 & f_1 \partial_y^2 f_2 \partial_y^2 h_1 h_2 & \partial_y f_1 \partial_y f_2 h_1 \partial_y^2 h_2 & \partial_y f_1 \partial_y f_2 \partial_y h_1 \partial_y h_2 \\
 & \partial_y f_1 \partial_y f_2 \partial_y^2 h_1 h_2 & \partial_y^2 f_1 f_2 h_1 \partial_y^2 h_2 & \partial_y^2 f_1 f_2 \partial_y h_1 \partial_y h_2 \\
 & \partial_y^2 f_1 f_2 \partial_y^2 h_1 h_2 & f_1 \partial_y^3 f_2 h_1 \partial_y h_2 & \partial_y f_1 \partial_y^2 f_2 h_1 \partial_y h_2 \\
 & \partial_y^2 f_1 \partial_y f_2 h_1 \partial_y h_2 & \partial_y^3 f_1 f_2 h_1 \partial_y h_2 & f_1 \partial_y^3 f_2 \partial_y h_1 h_2 \\
 & \partial_y f_1 \partial_y^2 f_2 \partial_y h_1 h_2 & \partial_y^2 f_1 \partial_y f_2 \partial_y h_1 h_2 & \partial_y^3 f_1 f_2 \partial_y h_1 h_2 \\
 & f_1 \partial_y^4 f_2 h_1 h_2 & \partial_y f_1 \partial_y^3 f_2 h_1 h_2 & \partial_y^2 f_1 \partial_y^2 f_2 h_1 h_2 \\
 & \partial_y^3 f_1 \partial_y f_2 h_1 h_2 & \partial_y^4 f_1 f_2 h_1 h_2\,. &
\end{array}
\tag{3.A.7}
$$

Let $f(y) \in C_4^{\text{pol}}(\mathcal{X})$ with ith derivative $\partial_y^i f(y)$ for $i = 1, \ldots, 4$. Suppose that Assumption 3.2.5 holds, and that the order of differentiation and integration may be exchanged when necessary. Then, for fixed $\lambda \in \Theta$,

$$
\begin{aligned}
&\mathcal{L}_\lambda f(y) \\
&\quad = a(y;\lambda)\partial_y f(y) + \tfrac{1}{2} b^2(y;\lambda)\partial_y^2 f(y) + \int_{\mathbb{R}} \left(f(y + c(y,z;\lambda)) - f(y)\right) \nu_\lambda(dz)\,,
\end{aligned}
\tag{3.A.8}
$$

$$
\begin{aligned}
&\partial_y \mathcal{L}_\lambda f(y) \\
&\quad = \partial_y a(y;\lambda)\partial_y f(y) + \left(a(y;\lambda) + \tfrac{1}{2}\partial_y b^2(y;\lambda)\right)\partial_y^2 f(y) + \tfrac{1}{2} b^2(y;\lambda)\partial_y^3 f(y) \\
&\qquad + \int_{\mathbb{R}} \left(\partial_y f(y + c(y,z;\lambda))\left(1 + \partial_y c(y,z;\lambda)\right) - \partial_y f(y)\right) \nu_\lambda(dz)\,,
\end{aligned}
\tag{3.A.9}
$$

$$
\begin{aligned}
&\partial_y^2 \mathcal{L}_\lambda f(y) \\
&\quad = \partial_y^2 a(y;\lambda)\partial_y f(y) + \left(2\partial_y a(y;\lambda) + \tfrac{1}{2}\partial_y^2 b^2(y;\lambda)\right)\partial_y^2 f(y) \\
&\qquad + \left(a(y;\lambda) + \partial_y b^2(y;\lambda)\right)\partial_y^3 f(y) + \tfrac{1}{2} b^2(y;\lambda)\partial_y^4 f(y) \\
&\qquad + \int_{\mathbb{R}} \left(\partial_y^2 f(y + c(y,z;\lambda))\left(1 + \partial_y c(y,z;\lambda)\right)^2 - \partial_y^2 f(y)\right) \nu_\lambda(dz) \\
&\qquad + \int_{\mathbb{R}} \partial_y f(y + c(y,z;\lambda))\partial_y^2 c(y,z;\lambda)\, \nu_\lambda(dz)\,,
\end{aligned} \tag{3.A.10}
$$

$$
\begin{aligned}
&\mathcal{L}_\lambda^2 f(y) \\
&\quad = a(y;\lambda)\partial_y \mathcal{L}_\lambda f(y) + \tfrac{1}{2} b^2(y;\lambda)\partial_y^2 \mathcal{L}_\lambda f(y) \\
&\qquad + \int_{\mathbb{R}} \left(\mathcal{L}_\lambda f(y + c(y,z;\lambda)) - \mathcal{L}_\lambda f(y)\right) \nu_\lambda(dz)\,.
\end{aligned} \tag{3.A.11}
$$

Now, using (3.A.5) when $f = f_1 f_2$, the preceding formulae may be rewritten as

$$
\begin{aligned}
&\mathcal{L}_\lambda f_1 f_2(y) \\
&\quad = a(y;\lambda)\left(f_1\partial_y f_2 + \partial_y f_1 f_2\right)(y) \\
&\qquad + \tfrac{1}{2} b^2(y;\lambda)\left(f_1\partial_y^2 f_2 + 2\partial_y f_1\partial_y f_2 + \partial_y^2 f_1 f_2\right)(y) \\
&\qquad + \int_{\mathbb{R}} \left(f_1 f_2(y + c(y,z;\lambda)) - f_1 f_2(y)\right) \nu_\lambda(dz)\,,
\end{aligned} \tag{3.A.12}
$$

$$
\begin{aligned}
&\partial_y \mathcal{L}_\lambda f_1 f_2(y) \\
&\quad = \partial_y a(y;\lambda)\left(f_1\partial_y f_2 + \partial_y f_1 f_2\right)(y) \\
&\qquad + \left(a(y;\lambda) + \tfrac{1}{2}\partial_y b^2(y;\lambda)\right)\left(f_1\partial_y^2 f_2 + 2\partial_y f_1\partial_y f_2 + \partial_y^2 f_1 f_2\right)(y) \\
&\qquad + \tfrac{1}{2} b^2(y;\lambda)\left(f_1\partial_y^3 f_2 + 3\partial_y f_1\partial_y^2 f_2 + 3\partial_y^2 f_1\partial_y f_2 + \partial_y^3 f_1 f_2\right)(y) \\
&\qquad + \int_{\mathbb{R}} \left(f_1\partial_y f_2 + \partial_y f_1 f_2\right)(y + c(y,z;\lambda))\left(1 + \partial_y c(y,z;\lambda)\right) \nu_\lambda(dz) \\
&\qquad - \int_{\mathbb{R}} \left(f_1\partial_y f_2 + \partial_y f_1 f_2\right)(y)\, \nu_\lambda(dz)\,,
\end{aligned} \tag{3.A.13}
$$

$$
\begin{aligned}
&\partial_y^2 \mathcal{L}_\lambda f_1 f_2(y) \\
&\quad = \partial_y^2 a(y;\lambda)\left(f_1\partial_y f_2 + \partial_y f_1 f_2\right)(y) \\
&\qquad + \left(2\partial_y a(y;\lambda) + \tfrac{1}{2}\partial_y^2 b^2(y;\lambda)\right)\left(f_1\partial_y^2 f_2 + 2\partial_y f_1\partial_y f_2 + \partial_y^2 f_1 f_2\right)(y) \\
&\qquad + \left(a(y;\lambda) + \partial_y b^2(y;\lambda)\right)\left(f_1\partial_y^3 f_2 + 3\partial_y f_1\partial_y^2 f_2 + 3\partial_y^2 f_1\partial_y f_2 + \partial_y^3 f_1 f_2\right)(y) \\
&\qquad + \tfrac{1}{2} b^2(y;\lambda)\left(f_1\partial_y^4 f_2 + 4\partial_y f_1\partial_y^3 f_2 + 6\partial_y^2 f_1\partial_y^2 f_2 + 4\partial_y^3 f_1\partial_y f_2 + \partial_y^4 f_1 f_2\right)(y) \\
&\qquad + \int_{\mathbb{R}} \left(f_1\partial_y^2 f_2 + 2\partial_y f_1\partial_y f_2 + \partial_y^2 f_1 f_2\right)(y + c(y,z;\lambda))\left(1 + \partial_y c(y,z;\lambda)\right)^2 \nu_\lambda(dz) \\
&\qquad - \int_{\mathbb{R}} \left(f_1\partial_y^2 f_2 + 2\partial_y f_1\partial_y f_2 + \partial_y^2 f_1 f_2\right)(y)\, \nu_\lambda(dz) \\
&\qquad + \int_{\mathbb{R}} \left(f_1\partial_y f_2 + \partial_y f_1 f_2\right)(y + c(y,z;\lambda))\partial_y^2 c(y,z;\lambda)\, \nu_\lambda(dz)\,,
\end{aligned} \tag{3.A.14}
$$

$$\begin{aligned}
&\mathcal{L}_\lambda^2 f_1 f_2(y) \\
&\quad = a(y;\lambda)\partial_y \mathcal{L}_\lambda f_1 f_2(y) + \tfrac{1}{2} b^2(y;\lambda)\partial_y^2 \mathcal{L}_\lambda f_1 f_2(y) \\
&\qquad + \int_{\mathbb{R}} \left(\mathcal{L}_\lambda f_1 f_2(y + c(y,z;\lambda)) - \mathcal{L}_\lambda f_1 f_2(y)\right) \nu_\lambda(dz)\,.
\end{aligned} \tag{3.A.15}$$

Condition 3.A.3 (For use with Assumption 3.4.8)**.** *Let $\tau_k(x, \mathbf{z}_k; \tilde{\alpha})$ be as defined in Definition 3.4.9. For all $\tilde{\alpha} \in A$ and $\theta \in \Theta$, it holds that*

$$\begin{aligned}
g_\beta(0, \tau_k(x, \mathbf{z}_k; \tilde{\alpha}), x; \theta) &= 0\,, \quad k = 1, 2 \\
\partial_y g_\beta(0, \tau_k(x, \mathbf{z}_k; \tilde{\alpha}), x; \theta) &= 0\,, \quad k = 0, 1
\end{aligned}$$

for all $x \in \mathcal{X}$ and $\tilde{\nu}$-almost all $\mathbf{z}_k \in \mathbb{R}^k$. ⋄

Condition 3.A.3 is a restatement of some of the conditions used to obtain rate optimality of G_n-estimators of a one-dimensional diffusion parameter β, see Conditions 3.4.10 and 3.4.14.

Lemma 3.A.4. *Suppose that Assumptions 3.2.5, 3.2.6, 3.4.8, and Condition 3.A.3 hold. Then, for $j_1 = 1, \ldots, d$ and $j_2 = d_1 + 1, \ldots, d$, the following formulae hold for all $x \in \mathcal{X}$ and $\theta \in \Theta$.*

$$\mathcal{L}_{\theta_0}\left(g_{j_1} g_{j_2}(0;\theta)\right)(x,x) = 0\,, \tag{3.A.16}$$

and, furthermore,

$$\begin{aligned}
&\mathcal{L}_{\theta_0}^2\left(g_{j_1} g_{j_2}(0;\theta)\right)(x,x) \\
&\quad = \tfrac{3}{2} b^2(x;\beta_0)\left(2a(x;\alpha_0) + \partial_y b^2(x;\beta_0)\right)\partial_y g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta) \\
&\qquad + \tfrac{1}{2} b^4(x;\beta_0)\left(2\partial_y g_{j_1}\partial_y^3 g_{j_2} + 3\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}\right)(0,x,x;\theta) \\
&\qquad + \int_{\mathbb{R}} \tfrac{1}{2}\left(b^2(\tau_1(x,z;\alpha_0);\beta_0) + b^2(x;\beta_0)\left(1 + \partial_y c(x,z;\alpha_0)\right)^2\right) \\
&\qquad\qquad \times g_{j_1}\partial_y^2 g_{j_2}(0,\tau_1(x,z;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz)\,,
\end{aligned} \tag{3.A.17}$$

$$\begin{aligned}
&g_{j_1}^{(1)}(x,x;\theta) \\
&\quad = -a(x;\alpha)\partial_y g_{j_1}(0,x,x;\theta) - \tfrac{1}{2} b^2(x;\beta)\partial_y^2 g_{j_1}(0,x,x;\theta) \\
&\qquad - \int_{\mathbb{R}} g_{j_1}(0,\tau_1(x,z;\alpha),x;\theta)\,\nu_\alpha(dz)\,,
\end{aligned} \tag{3.A.18}$$

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}(0;\theta) g_{j_2}^{(1)}(\theta)\right)(x,x) \\
&\quad = -\tfrac{1}{2} a(x;\alpha_0) b^2(x;\beta)\partial_y g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta) \\
&\qquad - \tfrac{1}{4} b^2(x;\beta) b^2(x;\beta_0)\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta) \\
&\qquad + b^2(x;\beta_0)\partial_y g_{j_1}(0,x,x;\theta)\partial_y g_{j_2}^{(1)}(x,x;\theta) \\
&\qquad + \int_{\mathbb{R}} g_{j_1}(0,\tau_1(x,z;\alpha_0),x;\theta) g_{j_2}^{(1)}(\tau_1(x,z;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz)\,,
\end{aligned} \tag{3.A.19}$$

and

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}^{(1)}(\theta)g_{j_2}(0;\theta)\right)(x,x)\\
&\quad= -\tfrac{1}{2}a(x;\alpha)b^2(x;\beta_0)\partial_y g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta)\\
&\qquad-\tfrac{1}{4}b^2(x;\beta)b^2(x;\beta_0)\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta)\\
&\qquad-\tfrac{1}{2}b^2(x;\beta_0)\left(\int_{\mathbb{R}} g_{j_1}(0,\tau_1(x,z;\alpha),x;\theta)\,\nu_\alpha(dz)\right)\partial_y^2 g_{j_2}(0,x,x;\theta)\,.
\end{aligned}\tag{3.A.20}$$

◇

Proof of Lemma 3.A.4. By Lemma 3.A.1, both expressions

$$\mathcal{L}_{\theta_0}(g_{j_1}g_{j_2}(0;\theta))(y,x) \quad\text{and}\quad \mathcal{L}_{\theta_0}^2(g_{j_1}g_{j_2}(0;\theta))(y,x)$$

are well-defined. Using (3.A.12),

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}(0;\theta)\right)(x,x)\\
&\quad= a(x;\alpha_0)\left(g_{j_1}\partial_y g_{j_2}+\partial_y g_{j_1}g_{j_2}\right)(0,x,x;\theta)\\
&\qquad+\tfrac{1}{2}b^2(x;\beta_0)\left(g_{j_1}\partial_y^2 g_{j_2}+2\partial_y g_{j_1}\partial_y g_{j_2}+\partial_y^2 g_{j_1}g_{j_2}\right)(0,x,x;\theta)\\
&\qquad+\int_{\mathbb{R}}\left(g_{j_1}g_{j_2}(0,\tau_1(x,z;\alpha_0),x;\theta)-g_{j_1}g_{j_2}(0,x,x;\theta)\right)\nu_{\alpha_0}(dz)\,,
\end{aligned}$$

and by Condition 3.A.3 and Lemma 3.2.9, (3.A.16) follows. Similarly,

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}(0;\theta)\right)(\tau_1(x,z_2;\alpha_0),x)\\
&\quad= a(\tau_1(x,z_2;\alpha_0);\alpha_0)\left(g_{j_1}\partial_y g_{j_2}+\partial_y g_{j_1}g_{j_2}\right)(0,\tau_1(x,z_2;\alpha_0),x;\theta)\\
&\qquad+\tfrac{1}{2}b^2(\tau_1(x,z_2;\alpha_0);\beta_0)\left(g_{j_1}\partial_y^2 g_{j_2}+2\partial_y g_{j_1}\partial_y g_{j_2}+\partial_y^2 g_{j_1}g_{j_2}\right)(0,\tau_1(x,z_2;\alpha_0),x;\theta)\\
&\qquad+\int_{\mathbb{R}} g_{j_1}g_{j_2}(0,\tau_2(x,\mathbf{z}_2;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz_1)\\
&\qquad-\int_{\mathbb{R}} g_{j_1}g_{j_2}(0,\tau_1(x,z_2;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz_1)\\
&\quad= \tfrac{1}{2}b^2(\tau_1(x,z_2;\alpha_0);\beta_0)g_{j_1}\partial_y^2 g_{j_2}(0,\tau_1(x,z_2;\alpha_0),x;\theta)
\end{aligned}\tag{3.A.21}$$

for $\tilde{\nu}$-almost all $z_2\in\mathbb{R}$. Using (3.A.13) and (3.A.14),

$$\partial_y\mathcal{L}_{\theta_0}g_{j_1}g_{j_2}(0,y,x;\theta)\Big|_{y=x}=\tfrac{3}{2}b^2(x;\beta_0)\partial_y g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta)\,. \tag{3.A.22}$$

and

$$\begin{aligned}
&\partial_y^2\mathcal{L}_{\theta_0}g_{j_1}g_{j_2}(0,y,x;\theta)\Big|_{y=x}\\
&\quad= 3\left(a(x;\alpha_0)+\partial_y b^2(x;\beta_0)\right)\partial_y g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta)\\
&\qquad+b^2(x;\beta_0)\left(2\partial_y g_{j_1}\partial_y^3 g_{j_2}+3\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}\right)(0,x,x;\theta)\\
&\qquad+\int_{\mathbb{R}}\left(1+\partial_y c(x,z;\alpha_0)\right)^2 g_{j_1}\partial_y^2 g_{j_2}(0,\tau_1(x,z;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz)\,.
\end{aligned}\tag{3.A.23}$$

Finally, by (3.A.15),

$$\begin{aligned}
&\mathcal{L}_{\theta_0}^2 \left(g_{j_1}g_{j_2}(0;\theta)\right)(x,x) \\
&\quad = a(x;\alpha_0)\ \partial_y \mathcal{L}_{\theta_0} g_{j_1}g_{j_2}(0,y,x;\theta)\Big|_{y=x} + \tfrac{1}{2}b^2(x;\beta_0)\ \partial_y^2 \mathcal{L}_{\theta_0} g_{j_1}g_{j_2}(0,y,x;\theta)\Big|_{y=x} \\
&\qquad + \int_{\mathbb{R}} \mathcal{L}_{\theta_0} g_{j_1}g_{j_2}(0,\tau_1(x,z;\alpha_0),x;\theta)\, \nu_{\alpha_0}(dz)\,.
\end{aligned}$$

Insert (3.A.21), (3.A.22) and (3.A.23) to obtain (3.A.17). Lemma 3.2.9 yields

$$g_j^{(1)}(x,x;\theta) = -\mathcal{L}_\theta \left(g_j(0;\theta)\right)(x,x)\,,$$

from which (3.A.18) follows. By (3.A.12) and (3.A.18), using that $\partial_y g_{j_2}(0,x,x;\theta) = 0$,

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}(0;\theta)g_{j_2}^{(1)}(\theta)\right)(x,x) \\
&\quad = \left(a(x;\alpha_0)\partial_y g_{j_1}(0,x,x;\theta) + \tfrac{1}{2}b^2(x;\beta_0)\partial_y^2 g_{j_1}(0,x,x;\theta)\right) g_{j_2}^{(1)}(x,x;\theta) \\
&\qquad + b^2(x;\beta_0)\partial_y g_{j_1}(0,x,x;\theta)\partial_y g_{j_2}^{(1)}(x,x;\theta) \\
&\qquad + \int_{\mathbb{R}} g_{j_1}(0,x+c(x,z;\alpha_0),x;\theta) g_{j_2}^{(1)}(x+c(x,z;\alpha_0),x;\theta)\, \nu_{\alpha_0}(dz) \\
&\quad = -\tfrac{1}{2}b^2(x;\beta)\left(a(x;\alpha_0)\partial_y g_{j_1}(0,x,x;\theta) + \tfrac{1}{2}b^2(x;\beta_0)\partial_y^2 g_{j_1}(0,x,x;\theta)\right)\partial_y^2 g_{j_2}(0,x,x;\theta) \\
&\qquad + b^2(x;\beta_0)\partial_y g_{j_1}(0,x,x;\theta)\partial_y g_{j_2}^{(1)}(x,x;\theta) \\
&\qquad + \int_{\mathbb{R}} g_{j_1}(0,\tau_1(x,z;\alpha_0),x;\theta) g_{j_2}^{(1)}(\tau_1(x,z;\alpha_0),x;\theta)\, \nu_{\alpha_0}(dz)
\end{aligned}$$

and

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}^{(1)}(\theta)g_{j_2}(0;\theta)\right)(x,x) \\
&\quad = \tfrac{1}{2}b^2(x;\beta_0)g_{j_1}^{(1)}(x,x;\theta)\partial_y^2 g_{j_2}(0,x,x;\theta) \\
&\quad = -\tfrac{1}{2}b^2(x;\beta_0)\left(a(x;\alpha)\partial_y g_{j_1}(0,x,x;\theta) + \tfrac{1}{2}b^2(x;\beta)\partial_y^2 g_{j_1}(0,x,x;\theta)\right)\partial_y^2 g_{j_2}(0,x,x;\theta) \\
&\qquad - \tfrac{1}{2}b^2(x;\beta_0)\left(\int_{\mathbb{R}} g_{j_1}(0,\tau_1(x,z;\alpha),x;\theta)\,\nu_\alpha(dz)\right)\partial_y^2 g_{j_2}(0,x,x;\theta)\,,
\end{aligned}$$

showing (3.A.19) and (3.A.20). □

Corollary 3.A.5. *Suppose that Assumptions 3.2.5, 3.2.6 and 3.4.8, and Condition 3.A.3 hold. Then, for $j_1, j_2 = d_1+1,\ldots,d$, it holds that for all $x \in \mathcal{X}$ and $\theta \in \Theta$,*

$$\mathcal{L}_{\theta_0}^2\left(g_{j_1}g_{j_2}(0;\theta)\right)(x,x) = \tfrac{3}{2}b^4(x;\beta_0)\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta)$$

and

$$\begin{aligned}
g_{j_1}^{(1)}(x,x;\theta) &= -\tfrac{1}{2}b^2(x;\beta)\partial_y^2 g_{j_1}(0,x,x;\theta)\,, \qquad (3.A.24)\\
\mathcal{L}_{\theta_0}\left(g_{j_1}(0;\theta)g_{j_2}^{(1)}(\theta)\right)(x,x) &= -\tfrac{1}{4}b^2(x;\beta)b^2(x;\beta_0)\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}(0,x,x;\theta)\,.
\end{aligned}$$

◇

Corollary 3.A.5 follows directly from Lemma 3.A.4.

Lemma 3.A.6. *Suppose that Assumptions 3.2.5, 3.2.6, 3.4.8, and Condition 3.A.3 hold. Then, for all $x \in \mathcal{X}$ and $\theta \in \Theta$, the following holds.*

(i) For $j_1, j_2, j_3 = 1, \ldots, d$ and $j_4 = d_1 + 1, \ldots, d$,

$$\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(x,x) = 0\,.$$

(ii) For $j_1, j_2 = 1, \ldots, d$ and $j_3, j_4 = d_1 + 1, \ldots, d$,

$$\begin{aligned}\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(0;\theta)\right)(x,x) &= 0 \\ \mathcal{L}_{\theta_0}\left(g^{(1)}_{j_1}(\theta)g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(x,x) &= 0 \\ \mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(0;\theta)g^{(1)}_{j_4}(\theta)\right)(x,x) &= 0\,.\end{aligned} \tag{3.A.25}$$

(iii) For $j_1 = 1, \ldots, d$ and $j_2, j_3, j_4 = d_1 + 1, \ldots, d$,

$$\mathcal{L}^2_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(x,x) = 0\,.$$

◇

Proof of Lemma 3.A.6. First, Lemma 3.A.6.(i) and (iii) are proven. By Lemma 3.A.1, the expressions $\mathcal{L}_{\theta_0}(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta))(y,x)$ and $\mathcal{L}^2_{\theta_0}(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta))(y,x)$ are well-defined for $x, y \in \mathcal{X}$. First, use (3.A.8) to write

$$\begin{aligned}&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(y,x) \\ &\quad = a(y;\alpha_0)\partial_y\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\right) + \tfrac{1}{2}b^2(y;\beta_0)\partial_y^2\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\right) \\ &\qquad + \int_{\mathbb{R}}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,\tau_1(y,z;\alpha_0),x;\theta) - g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\right)\nu_{\alpha_0}(dz)\end{aligned} \tag{3.A.26}$$

for $j_1, j_2, j_3 = 1, \ldots, d$ and $j_4 = d_1 + 1, \ldots, d$, and see that

$$\begin{aligned}&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(x,x) \\ &\quad = a(x;\theta_0)\partial_y\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,x,x;\theta)\right) + \tfrac{1}{2}b^2(x;\theta_0)\partial_y^2\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,x,x;\theta)\right)\,.\end{aligned}$$

Using the generalised Leibnitz formula, see (3.A.7), observe that all terms in the derivatives

$$\partial_y\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\right) \quad \text{and} \quad \partial_y^2\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\right)$$

contain at least one factor $g_j(0,y,x;\theta)$ for some $j = j_1, j_2, j_3, j_4$, meaning that

$$\partial_y\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,x,x;\theta)\right) = \partial_y^2\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,x,x;\theta)\right) = 0\,,$$

and Lemma 3.A.6.(i) follows.

Now, use (3.A.9) and (3.A.10) to write

$$\begin{aligned}
&\partial_y \mathcal{L}_{\theta_0} g_{j_1} g_{j_2} g_{j_3} g_{j_4}(0,y,x;\theta) \\
&\quad = \partial_y a(y;\alpha_0)\partial_y (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \left(a(y;\alpha_0) + \tfrac{1}{2}\partial_y b^2(y;\beta_0)\right)\partial_y^2 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \tfrac{1}{2} b^2(y;\beta_0)\partial_y^3 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \int_{\mathbb{R}} \partial_y (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,\tau_1(y,z;\alpha_0),x;\theta)\left(1 + \partial_y c(y,z;\alpha_0)\right)\nu_{\alpha_0}(dz) \\
&\qquad - \int_{\mathbb{R}} \partial_y (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta)\,\nu_{\alpha_0}(dz)
\end{aligned} \tag{3.A.27}$$

and

$$\begin{aligned}
&\partial_y^2 \left(\mathcal{L}_{\theta_0} g_{j_1} g_{j_2} g_{j_3} g_{j_4}(0,y,x;\theta)\right) \\
&\quad = \partial_y^2 a(y;\alpha_0)\partial_y (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \left(2\partial_y a(y;\alpha_0) + \tfrac{1}{2}\partial_y^2 b^2(y;\beta_0)\right)\partial_y^2 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \left(a(y;\alpha_0) + \partial_y b^2(y;\beta_0)\right)\partial_y^3 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \tfrac{1}{2} b^2(y;\beta_0)\partial_y^4 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta) \\
&\qquad + \int_{\mathbb{R}} \partial_y^2 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,\tau_1(y,z;\alpha_0),x;\theta)\left(1 + \partial_y c(y,z;\alpha_0)\right)^2 \nu_{\alpha_0}(dz) \\
&\qquad + \int_{\mathbb{R}} \partial_y (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,\tau_1(y,z;\alpha_0),x;\theta)\partial_y^2 c(y,z;\alpha_0)\,\nu_{\alpha_0}(dz) \\
&\qquad - \int_{\mathbb{R}} \partial_y^2 (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta)\,\nu_{\alpha_0}(dz)\,.
\end{aligned} \tag{3.A.28}$$

Using the generalised Leibnitz formula again, see (3.A.7), it is seen that all terms in the derivatives

$$\partial_y^i (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta)\,, \quad i = 1,2\,,$$

contain at least one factor $g_j(0,y,x;\theta)$ for some $j = j_2, j_3, j_4$, and all terms in the derivatives

$$\partial_y^i (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,y,x;\theta)\,, \quad i = 3,4\,,$$

contain at least one factor $g_j(0,y,x;\theta)$ or $\partial_y g_j(0,y,x;\theta)$ for some $j = j_2, j_3, j_4$. So, for $j_1 = 1,\ldots,d$ and $j_2, j_3, j_4 = d_1+1,\ldots,d$,

$$\begin{aligned}
\partial_y^i (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,\tau_1(x,z;\alpha_0),x;\theta) &= 0\,, \quad i = 1,2 \\
\partial_y^i (g_{j_1} g_{j_2} g_{j_3} g_{j_4})(0,x,x;\theta) &= 0\,, \quad i = 1,2,3,4\,.
\end{aligned} \tag{3.A.29}$$

Inserting into (3.A.27) and (3.A.28), it follows that

$$\partial_y \mathcal{L}_{\theta_0} g_{j_1} g_{j_2} g_{j_3} g_{j_4}(0,y,x;\theta)\Big|_{y=x} = \partial_y^2 \mathcal{L}_{\theta_0} g_{j_1} g_{j_2} g_{j_3} g_{j_4}(0,y,x;\theta)\Big|_{y=x} = 0\,.$$

Furthermore, by (3.A.26), (3.A.29) and Condition 3.A.3,

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(\tau_1(x,z_2;\alpha_0),x)\\
&\quad= a(x+c(x,z;\alpha_0);\alpha_0)\partial_y(g_{j_1}g_{j_2}g_{j_3}g_{j_4})(0,\tau_1(x,z_2;\alpha_0),x;\theta)\\
&\qquad+\tfrac{1}{2}b^2(\tau_1(x,z;\alpha_0);\beta_0)\partial_y^2(g_{j_1}g_{j_2}g_{j_3}g_{j_4})(0,\tau_1(x,z_2;\alpha_0),x;\theta)\\
&\qquad+\int_{\mathbb{R}} g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,\tau_2(x,\mathbf{z}_2;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz_1)\\
&\qquad-\int_{\mathbb{R}} g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,\tau_1(x,z_2;\alpha_0),x;\theta)\,\nu_{\alpha_0}(dz_1)\\
&\quad=0
\end{aligned}$$

for $\tilde{\nu}$-almost all $z_2 \in \mathbb{R}$. By (3.A.11),

$$\begin{aligned}
&\mathcal{L}_{\theta_0}^2\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(x,x)\\
&\quad= a(x;\alpha_0)\;\partial_y\mathcal{L}_{\theta_0}g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\Big|_{y=x}\\
&\qquad+\tfrac{1}{2}b^2(x;\beta_0)\;\partial_y^2\mathcal{L}_{\theta_0}g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0,y,x;\theta)\Big|_{y=x}\\
&\qquad+\int_{\mathbb{R}}\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(\tau_1(x,z;\alpha_0),x)\,\nu_{\alpha_0}(dz)\\
&\qquad-\int_{\mathbb{R}}\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(x,x)\,\nu_{\alpha_0}(dz)
\end{aligned}$$

and Lemma 3.A.6.(iii) follows.

In order to prove Lemma 3.A.6.(ii) for $j_1, j_2 = 1,\ldots,d$ and $j_3, j_4 = d_1+1,\ldots,d$, recall that by Condition 3.A.3, $g_{j_3}(0,\tau_1(x,z;\alpha_0),x;\theta) = 0$ for $\tilde{\nu}$-almost all z. First write

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(0;\theta)\right)(y,x)\\
&\quad= a(y;\alpha_0)\partial_y(g_{j_1}g_{j_2}g_{j_3})(0,y,x;\theta)+\tfrac{1}{2}b^2(y;\beta_0)\partial_y^2(g_{j_1}g_{j_2}g_{j_3})(0,y,x;\theta)\\
&\qquad+\int_{\mathbb{R}}\left(g_{j_1}g_{j_2}g_{j_3}(0,y+c(y,z;\alpha_0),x;\theta)-g_{j_1}g_{j_2}g_{j_3}(0,y,x;\theta)\right)\nu_{\alpha_0}(dz)\,.
\end{aligned}$$

By (3.A.6) it is seen that each term in

$$\partial_y(g_{j_1}g_{j_2}g_{j_3})(0,y,x;\theta)\quad\text{and}\quad\partial_y^2(g_{j_1}g_{j_2}g_{j_3})(0,y,x;\theta)$$

contains at least one factor $g_j(0,y,x;\theta)$ for some $j = j_1, j_2, j_3$, so

$$\partial_y(g_{j_1}g_{j_2}g_{j_3})(0,x,x;\theta)=\partial_y^2(g_{j_1}g_{j_2}g_{j_3})(0,x,x;\theta)=0$$

and (3.A.25) follows. Now, write

$$\begin{aligned}
&\mathcal{L}_{\theta_0}\left(g_i g_{j_2}g_{j_3}(0;\theta)g_j^{(1)}(\theta)\right)(y,x)\\
&\quad= a(y;\alpha_0)\partial_y\left(g_i g_{j_2}g_{j_3}(0,y,x;\theta)g_j^{(1)}(y,x;\theta)\right)\\
&\qquad+\tfrac{1}{2}b^2(y;\beta_0)\partial_y^2\left(g_i g_{j_2}g_{j_3}(0,y,x;\theta)g_j^{(1)}(y,x;\theta)\right)
\end{aligned}$$

$$+ \int_{\mathbb{R}} g_i g_{j_2} g_{j_3}(0, y + c(y, z; \alpha_0), x; \theta) g_j^{(1)}(y + c(y, z; \alpha_0), x; \theta)\, \nu_{\alpha_0}(dz)$$
$$- \int_{\mathbb{R}} g_i g_{j_2} g_{j_3}(0, y, x; \theta) g_j^{(1)}(y, x; \theta)\, \nu_{\alpha_0}(dz)\,.$$

where $(i, j) = (j_1, j_4)$ or $(i, j) = (j_4, j_1)$. In either case, each term in

$$\partial_y \left(g_i g_{j_2} g_{j_3}(0, y, x; \theta) g_j^{(1)}(y, x; \theta) \right) \quad \text{and} \quad \partial_y^2 \left(g_i g_{j_2} g_{j_3}(0, y, x; \theta) g_j^{(1)}(y, x; \theta) \right)$$

contains a factor $g_k(0, y, x; \theta)$ for at least one of $k = j_1, j_2, j_3, j_4$, so

$$\partial_y \left(g_i g_{j_2} g_{j_3}(0, x, x; \theta) g_j^{(1)}(x, x; \theta) \right) = \partial_y^2 \left(g_i g_{j_2} g_{j_3}(0, x, x; \theta) g_j^{(1)}(x, x; \theta) \right) = 0\,,$$

and the remaining results of Lemma 3.A.6.(ii) follow. □

Lemma 3.A.7. *Suppose that Assumptions 3.2.5, 3.2.6, 3.4.8, and Condition 3.A.3 hold, and that*

$$\partial_y^2 \partial_\alpha g_\beta(0, x, x; \theta) = 0 \tag{3.A.30}$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Then

$$\begin{aligned} \mathcal{L}_{\theta_0} \left(\partial_\alpha g_\beta(0; \theta) \right)(x, x) &= 0 \\ \partial_\alpha \mathcal{L}_\theta \left(g_\beta(0; \theta) \right)(x, x) &= 0 \\ \partial_\alpha g_\beta^{(1)}(x, x; \theta) &= 0 \\ \mathcal{L}_{\theta_0} \left((\partial_\alpha g_\beta)^2(0; \theta) \right)(x, x) &= 0 \\ \mathcal{L}_{\theta_0}^2 \left((\partial_\alpha g_\beta)^2(0; \theta) \right)(x, x) &= 0\,, \end{aligned}$$

and, for $j = d_1 + 1, \ldots, d$, $k = 1, \ldots, d_1$,

$$\mathcal{L}_{\theta_0} \left(\partial_{\theta_k} g_j(0; \theta) \partial_{\theta_k} g_j^{(1)}(\theta) \right)(x, x) = 0\,.$$

◇

Proof of Lemma 3.A.7. Observe first that by Lemma 3.2.9, Condition 3.A.3 and the condition (3.A.30), also

$$\begin{aligned} \partial_\alpha g_\beta(0, \tau_m(x, \mathbf{z}_m; \alpha_0), x; \theta) &= 0\,, \quad m = 0, 1, 2\,, \\ \partial_y \partial_\alpha g_\beta(0, \tau_m(x, \mathbf{z}_m; \alpha_0), x; \theta) &= 0\,, \quad m = 0, 1\,, \end{aligned}$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Now, using Remark 3.2.10,

$$\begin{aligned} &\mathcal{L}_{\theta_0} \left(\partial_\alpha g_\beta(0; \theta) \right)(x, x) \\ &\quad = a(x; \alpha_0) \partial_y \partial_\alpha g_\beta(0, x, x; \theta) + \tfrac{1}{2} b^2(x; \beta_0) \partial_y^2 \partial_\alpha g_\beta(0, x, x; \theta) \\ &\qquad + \int_{\mathbb{R}} \partial_\alpha g_\beta(0, x + c(x, z; \alpha_0), x; \theta) \nu_{\alpha_0}(dz) \\ &\quad = 0 \end{aligned}$$

and

$$\begin{aligned}
&\partial_\alpha \mathcal{L}_\theta \left(g_\beta(0;\theta)\right)(x,x) \\
&\quad = \mathcal{L}_\theta(\partial_\alpha g_\beta(0,\theta))(x,x) + \partial_y g_\beta(0,x,x;\theta)\partial_\alpha a(x;\alpha) + \tfrac{1}{2}\partial_y^2 g_\beta(0,x,x;\theta)\partial_\alpha b^2(x;\beta) \\
&\qquad + \int_{\mathbb{R}} \partial_y g_\beta(0, x + c(x,z;\alpha), x;\theta)\partial_\alpha c(x,z;\alpha)\, \nu_\alpha(dz) \\
&\qquad + \int_{\mathbb{R}} g_\beta(0, x + c(x,z;\alpha), x;\theta)\partial_\alpha q(z;\alpha)\, \tilde{\nu}(dz) \\
&\quad = 0\,,
\end{aligned}$$

and by (3.A.24),

$$\partial_\alpha g_\beta^{(1)}(x,x;\theta) = -\tfrac{1}{2} b^2(x;\beta)\partial_y^2 \partial_\alpha g_\beta(0,x,x;\theta) = 0\,,$$

proving the first three equalities.

Now, let $j = d_1 + 1, \dots, d$ and $k = 1, \dots, d_1$. By (3.A.12), (3.A.13) and (3.A.14), as

$$\begin{aligned}
\partial_y^i \partial_{\theta_k} g_j(0,x,x;\theta) &= 0\,, \quad i = 0,1,2 \\
\partial_y^i \partial_{\theta_k} g_j(0, \tau_1(x;z_1;\alpha_0), x;\theta) &= 0\,, \quad i = 0,1\,,
\end{aligned} \tag{3.A.31}$$

it holds that

$$\begin{aligned}
\mathcal{L}_{\theta_0}\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(x,x) &= 0 \\
\partial_y \mathcal{L}_{\theta_0}\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(y,x)\Big|_{y=x} &= 0 \\
\partial_y^2 \mathcal{L}_{\theta_0}\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(y,x)\Big|_{y=x} &= 0\,.
\end{aligned}$$

Since also

$$\partial_{\theta_k} g_j(0, \tau_2(x;\mathbf{z}_2;\alpha_0), x;\theta) = 0\,,$$

it holds that

$$\mathcal{L}_{\theta_0}\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(x + c(x,z;\alpha_0), x) = 0$$

as well, and it follows from (3.A.15) that

$$\mathcal{L}_{\theta_0}^2\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(x,x) = 0\,.$$

Finally, by (3.A.31) and (3.A.12),

$$\mathcal{L}_{\theta_0}\left(\partial_{\theta_k} g_j(0;\theta)\partial_{\theta_k} g_j^{(1)}(\theta)\right)(x,x) = 0,$$

thus completing the proof of the last three equalities. □

Lemma 3.A.8. *Let $\lambda \in \Theta$ be given, and suppose that Assumptions 3.2.5 and 3.2.6 hold. Then,* *(i) for* $j, j_1, j_2 = 1, \ldots, d,$

$$\begin{aligned}&(x;\theta) \mapsto \mathcal{L}_\lambda(g_j(0,\theta))(x,x)\\&(x;\theta) \mapsto \mathcal{L}_\theta(g_j(0,\theta))(x,x)\\&(x;\theta) \mapsto \mathcal{L}_\lambda(g_{j_1}g_{j_2}(0,\theta))(x,x)\end{aligned}$$

are $C^{pol}_{1,2}(\mathcal{X}\times\Theta)$ *functions.*

(ii) under the additional Assumption 3.4.8, for $j_1 = 1, \ldots, d$ *and* $j_2 = d_1 + 1, \ldots, d,$

$$\begin{aligned}&(x;\theta) \mapsto \mathcal{L}^2_\lambda\left(g_{j_1}g_{j_2}(0;\theta)\right)(x,x)\\&(x;\theta) \mapsto \mathcal{L}_\lambda\left(g_{j_1}(0;\theta)g^{(1)}_{j_2}(\theta)\right)(x,x)\\&(x;\theta) \mapsto \mathcal{L}_\lambda\left(g^{(1)}_{j_1}(\theta)g_{j_2}(0;\theta)\right)(x,x)\end{aligned}$$

are $C^{pol}_{1,1}(\mathcal{X}\times\Theta)$*-functions.*

$\diamond$

Proof of Lemma 3.A.8. Note first that if $f(y,x;\theta) \in C^{\text{pol}}_{1,1,s}(\mathcal{X}^2\times\Theta)$, then it holds that $\tilde{f}(x;\theta) = f(x,x;\theta) \in C^{\text{pol}}_{1,s}(\mathcal{X}\times\Theta)$. Using Lemmas 3.A.1 and 3.A.2, it is seen that

$$\begin{aligned}&(x;\theta) \mapsto \int_{\mathbb{R}} g_j(0, x + c(x,z;\theta_0), x;\theta)\,\nu_{\theta_0}(dz)\\&(x;\theta) \mapsto \int_{\mathbb{R}} g_{j_1}g_{j_2}(0, x + c(x,z;\theta_0), x;\theta)\,\nu_{\theta_0}(dz)\\&(x;\theta) \mapsto \int_{\mathbb{R}} g_j(0, x + c(x,z;\theta), x;\theta)\,\nu_{\theta}(dz)\end{aligned}$$

are $C^{\text{pol}}_{1,2}(\mathcal{X}\times\Theta)$ functions. Then, Lemma 3.A.8.(i) follows from the expressions in Remark 3.2.10.

By (3.A.12) and Lemma 3.A.1, under the additional Assumption 3.4.8,

$$\begin{aligned}&\mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}(0;\theta)\right)(y,x)\\&\quad= a(y;\alpha_0)\left(g_{j_1}\partial_y g_{j_2} + \partial_y g_{j_1}g_{j_2}\right)(0,y,x;\theta)\\&\qquad+ \tfrac{1}{2}b^2(y;\beta_0)\left(g_{j_1}\partial^2_y g_{j_2} + \partial_y g_{j_1}\partial_y g_{j_2} + \partial^2_y g_{j_1}g_{j_2}\right)(0,y,x;\theta)\\&\qquad+ \int_{\mathbb{R}}\left(g_{j_1}g_{j_2}(0, y + c(y,z;\alpha_0), x;\theta) - g_{j_1}g_{j_2}(0,y,x;\theta)\right)\nu_{\alpha_0}(dz)\end{aligned}$$

is a $C^{\text{pol}}_{4,1,2}(\mathcal{X}^2\times\Theta)$ function, so by (3.A.15), $\mathcal{L}^2_{\theta_0}(g_{j_1}g_{j_2}(0;\theta))(x,x)$ is a $C^{\text{pol}}_{1,2}(\mathcal{X}^2\times\Theta)$ function. Finally, by (3.A.19), (3.A.20) and Lemma 3.A.1, the remaining results of Lemma 3.A.8.(ii) follow as well. $\square$

3.A.3 (Conditional) Expectation Inequalities

Lemma 3.A.9. *Let $\lambda \in \Theta$ be given. Suppose that Assumption 3.2.5 holds, and that $f(t,y,x;\theta) \in C^{pol}_{1,2,0,0}((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ Then, for $0 \le s \le t \le s + \Delta_0$,*

$$
\begin{aligned}
&f(t-s, X_t, X_s;\theta) - f(s, X_s, X_s;\theta) \\
&\quad = \int_s^t f_1(u-s, X_{u-}, X_s;\theta)\, du + \int_s^t f_2(u-s, X_{u-}, X_s;\theta)\, dW_u \\
&\qquad + \int_s^t \int_{\mathbb{R}} f_3(u-s, X_{u-}, X_s, z;\theta)\, (N^\lambda - \mu_\lambda)(du, dz)\,,
\end{aligned}
$$

under $\mathbb{P}_\lambda$, where f_1, f_2 and f_3 are given by

$$
\begin{aligned}
f_1(t,y,x;\theta) &= \partial_t f(t,y,x;\theta) + \mathcal{L}_\lambda(f(t;\theta))(y,x) \\
f_2(t,y,x;\theta) &= b(y;\lambda)\partial_y f(t,y,x;\theta) \\
f_3(t,y,x,z;\theta) &= f(t, y + c(y,z;\lambda), x;\theta) - f(t,y,x;\theta)\,,
\end{aligned}
$$

and where $\mathbf{M}^{(1)} = (M^{(1)}_v)_{v\ge 0}$, $\mathbf{M}^{(2)} = (M^{(2)}_v)_{v\ge 0}$ given by

$$
\begin{aligned}
M^{(1)}_v &= \int_0^v 1_{(s,t]}(u) f_2(u-s, X_{u-}, X_s;\theta)\, dW_u \\
M^{(2)}_v &= \int_0^v \int_{\mathbb{R}} 1_{(s,t]}(u) f_3(u-s, X_{u-}, X_s, z;\theta)\, (N^\lambda - \mu_\lambda)(du, dz)
\end{aligned}
$$

are $(\mathcal{F}_v)_{v\ge 0}$-martingales. ◇

Lemma 3.A.9 is essentially Itô's formula for stochastic differential equations with jumps of the form (3.2.1), see Applebaum (2009, Chapter 4.4.2). Assumption 3.2.5 and Lemma 3.A.1 ensure the martingale properties of the stochastic integrals by Applebaum (2009, Theorem 4.3.2).

Assumption 3.A.10. *Let $f(t,y,x;\theta) \in C^{pol}_{1,2,0,0}((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ with $f(0,x,x;\theta) = 0$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$.* ◇

Lemma 3.A.11. *Suppose that Assumption 3.2.5 holds, and that $f(t,y,x;\theta)$ satisfies Assumption 3.A.10. Let $p = 2^q$ for some $q \in \mathbb{N}$, and write*

$$
\begin{aligned}
f_1(t,y,x;\theta) &= \partial_t f(t,y,x;\theta) + \mathcal{L}_{\theta_0}\left(f(t;\theta)\right)(y,x) \\
f_2(t,y,x;\theta) &= b(y;\theta_0)\partial_y f(t,y,x;\theta) \\
f_3(t,y,x,z;\theta) &= f(t, y + c(y,z;\theta_0), x;\theta) - f(t,y,x;\theta)\,.
\end{aligned}
\tag{3.A.32}
$$

For any function $h(\,\cdot\,;\theta)$, let $Dh(\,\cdot\,;\theta,\theta') = h(\,\cdot\,;\theta) - h(\,\cdot\,;\theta')$. Then, there exist constants

$C_p > 0$ *such that*

$$\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^p\right)$$
$$\leq (n\Delta_n)^{p-1} C_p \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(Df_1(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}; \theta, \theta')^p\right) du$$
$$+ (n\Delta_n)^{p/2-1} C_p \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(Df_2(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}; \theta, \theta')^p\right) du$$
$$+ \sum_{l=1}^{q} (n\Delta_n)^{2^{q-l}-1} C_p \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(Df_3(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}, z; \theta, \theta')^p\right) \nu_{\theta_0}(dz)\, du$$

for all $\theta, \theta' \in \Theta$ *and* $n \in \mathbb{N}$. ◇

Proof of Lemma 3.A.11. By Ito's formula (Lemma 3.A.9),

$$\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^p\right)$$
$$\leq C_p \mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} Df_1(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')\, du\right|^p\right) \tag{3.A.33}$$
$$+ C_p \mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} Df_2(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')\, dW_u\right|^p\right) \tag{3.A.34}$$
$$+ C_p \mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} Df_3(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z; \theta, \theta')\, (N^{\theta_0} - \mu_{\theta_0})(du, dz)\right|^p\right) \tag{3.A.35}$$

for suitable constants $C_p > 0$. Starting with (3.A.33), and using Jensen's inequality twice,

$$\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} Df_1(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')\, du\right|^p\right)$$
$$= (n\Delta_n)^p \mathbb{E}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} Df_1(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')\, du\right|^p\right)$$
$$\leq (n\Delta_n)^{p-1} \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(Df_1(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')^p\right) du \tag{3.A.36}$$

Now, consider (3.A.34). By the martingale properties of the stochastic integral, the Burkholder-Davis-Gundy inequality may be used to deduce the existence of a constant $C_p > 0$ such that

$$\mathbb{E}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\Delta_n} \int_{t_{i-1}^n}^{t_i^n} Df_2(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')\, dW_u\right|^p\right)$$
$$\leq C_p \mathbb{E}_{\theta_0}\left(\left|\frac{1}{n^2}\sum_{i=1}^{n} \frac{1}{\Delta_n^2} \int_{t_{i-1}^n}^{t_i^n} Df_2(u - t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}; \theta, \theta')^2\, du\right|^{p/2}\right).$$

Now, in the same manner as before, we may write

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n} Df_2(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n};\theta,\theta')\,dW_u\right|^p\right)\\
&= (n\Delta_n)^p\mathbb{E}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Delta_n}\int_{t_{i-1}^n}^{t_i^n} Df_2(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n};\theta,\theta')\,dW_u\right|^p\right)\\
&\leq (n\Delta_n)^p C_p\mathbb{E}_{\theta_0}\left(\left|\frac{1}{n^2}\sum_{i=1}^{n}\frac{1}{\Delta_n^2}\int_{t_{i-1}^n}^{t_i^n} Df_2(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n};\theta,\theta')^2\,du\right|^{p/2}\right)\\
&\leq (n\Delta_n)^{p/2} C_p\mathbb{E}_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Delta_n}\int_{t_{i-1}^n}^{t_i^n} Df_2(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n};\theta,\theta')^2\,du\right|^{p/2}\right)\\
&\leq (n\Delta_n)^{p/2-1} C_p\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\mathbb{E}_{\theta_0}\left(Df_2(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n};\theta,\theta')^p\right)du\,. \qquad (3.A.37)
\end{aligned}
$$

Finally, for (3.A.35), let $\mathbf{M}^{(k)} = (M_v^{(k)})_{v\geq 0}$ and $\mathbf{S}^{(k)} = (S_v^{(k)})_{v\geq 0}$ be given by

$$
\begin{aligned}
M_v^{(k)} &= \int_0^v\int_{\mathbb{R}}\sum_{i=1}^{n}\mathbf{1}_{(t_{i-1}^n,t_i^n]}(u)Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^k(N^{\theta_0}-\mu_{\theta_0})(du,dz)\\
S_v^{(k)} &= \int_0^v\int_{\mathbb{R}}\sum_{i=1}^{n}\mathbf{1}_{(t_{i-1}^n,t_i^n]}(u)Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^k\,\nu_{\theta_0}(dz)\,du
\end{aligned}
$$

for $k\in\mathbb{N}$, and note that the quadratic variation of $\mathbf{M}^{(k)}$ may be written as

$$
\begin{aligned}
[\mathbf{M}^{(k)},\mathbf{M}^{(k)}]_v &= \int_0^v\int_{\mathbb{R}}\sum_{i=1}^{n}\mathbf{1}_{(t_{i-1}^n,t_i^n]}(u)Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^{2k}\,N^{\theta_0}(du,dz)\\
&= M_v^{(2k)} + S_v^{(2k)}\,.
\end{aligned}
$$

$\mathbf{M}^{(k)}$ is an $(\mathcal{F}_v)_{v\geq 0}$ martingale, so by the Burkholder-Davis-Gundy inequality, it holds that for any $m\geq 1$, there exist constants $C_m>0$ such that

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(|M_v^{(k)}|^m\right)\\
&\quad\leq C_m\mathbb{E}_{\theta_0}\left([\mathbf{M}^{(k)},\mathbf{M}^{(k)}]_v^{m/2}\right)\\
&\quad\leq C_m\mathbb{E}_{\theta_0}\left(\left(M_v^{(2k)}\right)^{m/2}\right) + C_m\mathbb{E}_{\theta_0}\left(\left(S_v^{(2k)}\right)^{m/2}\right).
\end{aligned}
$$

In particular, inserting 2^j in place of k and 2^{q-j} in place of m for $j\in\{0,1,\ldots,q-1\}$,

$$
\mathbb{E}_{\theta_0}\left(\left(M_v^{(2^j)}\right)^{2^{q-j}}\right)\leq C_p\mathbb{E}_{\theta_0}\left(\left(M_v^{(2^{j+1})}\right)^{2^{q-(j+1)}}\right) + C_p\mathbb{E}_{\theta_0}\left(\left(S_v^{(2^{j+1})}\right)^{2^{q-(j+1)}}\right).
$$

This inequality may be used iteratively to obtain

$$
\mathbb{E}_{\theta_0}\left(\left(M_v^{(1)}\right)^p\right)\leq C_p\mathbb{E}_{\theta_0}\left(M_v^{(p)}\right) + C_p\sum_{l=1}^{q}\mathbb{E}_{\theta_0}\left(\left(S_v^{(2^l)}\right)^{2^{q-l}}\right),
$$

and since $\mathbb{E}_{\theta_0}(M_v^{(p)}) = 0$ by properties of the Poisson integral,

$$\mathbb{E}_{\theta_0}\left(\left(M_v^{(1)}\right)^p\right) \leq C_p \sum_{l=1}^{q} \mathbb{E}_{\theta_0}\left(\left(S_v^{(2^l)}\right)^{2^{q-l}}\right).$$

That is,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}} Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')\,(N^{\theta_0}-\mu_{\theta_0})(du,dz)\right|^p\right)\\
&\quad\leq C_p\sum_{l=1}^{q}\mathbb{E}_{\theta_0}\left(\left(\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}} Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^{2^l}\,\nu_{\theta_0}(dz)\,du\right)^{2^{q-l}}\right)
\end{aligned}$$

Recalling that ν_θ has density $q(\cdot\,;\theta)$ with respect to $\tilde{\nu}$, where $q(z;\theta) = \xi(\theta)p(z;\theta)$ and $p(\cdot\,;\theta)$ is a probability density with respect to $\tilde{\nu}$, Jensen's inequality is used twice to write

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left(\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}} Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^{2^l}\,\nu_{\theta_0}(dz)\,du\right)^{2^{q-l}}\right)\\
&= (\xi(\theta_0)n\Delta_n)^{2^{q-l}}\\
&\qquad\times\mathbb{E}_{\theta_0}\left(\left(\frac{1}{n}\sum_{i=1}^{n}\frac{1}{\Delta_n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}} Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^{2^l}\,p(z;\theta_0)\,\tilde{\nu}(dz)\,du\right)^{2^{q-l}}\right)\\
&\leq (\xi(\theta_0)n\Delta_n)^{2^{q-l}-1}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}}\mathbb{E}_{\theta_0}\left(Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^p\right)\nu_{\theta_0}(dz)\,du\\
&= (n\Delta_n)^{2^{q-l}-1}C_p\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}}\mathbb{E}_{\theta_0}\left(Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')^p\right)\nu_{\theta_0}(dz)\,du
\end{aligned}$$

Thus,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}} Df_3(u-t_{i-1}^n, X_{u-}, X_{t_{i-1}^n}, z;\theta,\theta')\,(N^{\theta_0}-\mu_{\theta_0})(du,dz)\right|^p\right)\\
&\leq\sum_{l=1}^{q}(n\Delta_n)^{2^{q-l}-1}C_p\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}}\mathbb{E}_{\theta_0}\left(Df_3(u-t_{i-1}^n, X_u, X_{t_{i-1}^n}, z;\theta,\theta')^p\right)\nu_{\theta_0}(dz)\,du. \quad (3.A.38)
\end{aligned}$$

Note that for fixed $\omega \in \Omega$, $X_u(\omega) \neq X_{u-}(\omega)$ for at most countably many u in any finite interval $I \subseteq [0,\infty)$. Tonelli's theorem for non-negative functions was therefore used on the right-hand side of (3.A.38) to exchange the integration order and see that the X_{u-} could be replaced by X_u in the Lebesgue integral.

Now, inserting equations (3.A.36), (3.A.37) and (3.A.38) instead of (3.A.33), (3.A.34) and (3.A.35) the desired result follows. □

Lemma 3.A.12. *Suppose that Assumption 3.2.5 holds and let $m \in \mathbb{N}_0$. Define $\mathbf{z}_m = (z_1,\ldots,z_m)$, with the convention $\mathbf{z}_0 = ()$, and assume that $(t,y,x;\theta) \mapsto f(t,y,x,\mathbf{z}_m;\theta)$*

satisfies Assumption 3.A.10 for $\tilde{\nu}$-almost all $\mathbf{z}_m \in \mathbb{R}^m$. Let $p = 2^q$ for some $q \in \mathbb{N}$, and define

$$\begin{aligned} f_1(t, y, x, \mathbf{z}_m; \theta) &= \partial_t f(t, y, x, \mathbf{z}_m; \theta) + \mathcal{L}_{\theta_0} \left(f(t, \mathbf{z}_m; \theta) \right)(y, x) \\ f_2(t, y, x, \mathbf{z}_m; \theta) &= b(y; \theta_0) \partial_y f(t, y, x, \mathbf{z}_m; \theta) \\ f_3(t, y, x, \mathbf{z}_m, z; \theta) &= f(t, y + c(y, z; \theta_0), x, \mathbf{z}_m; \theta) - f(t, y, x, \mathbf{z}_m; \theta) \,. \end{aligned}$$

For any function $h(\,\cdot\,; \theta)$, let $Dh(\,\cdot\,; \theta, \theta') = h(\,\cdot\,; \theta) - h(\,\cdot\,; \theta')$. Then, there exist constants $C_p > 0$ such that

$$\begin{aligned} &\mathbb{E}_{\theta_0} \left(Df(t - s, X_t, X_s, \mathbf{z}_m; \theta, \theta')^p \right) \\ &\quad \leq (t - s)^{p-1} C_p \int_s^t \mathbb{E}_{\theta_0} \left(Df_1(u - s, X_u, X_s, \mathbf{z}_m; \theta, \theta')^p \right) du \\ &\qquad + (t - s)^{p/2-1} C_p \int_s^t \mathbb{E}_{\theta_0} \left(Df_2(u - s, X_u, X_s, \mathbf{z}_m; \theta, \theta')^p \right) du \\ &\qquad + \left(\sum_{l=1}^{q} (t - s)^{2^{q-l}-1} \right) C_p \int_s^t \int_{\mathbb{R}} \mathbb{E}_{\theta_0} \left(Df_3(u - s, X_u, X_s, \mathbf{z}_m, z; \theta, \theta')^p \right) \nu_{\theta_0}(dz) \, du \end{aligned}$$

for all $\theta, \theta' \in \Theta$, $0 \leq s < t \leq s + \Delta_0$, and $\tilde{\nu}$-almost all $\mathbf{z}_m$. ◇

The proof of Lemma 3.A.12 is identical to the proof of Lemma 3.A.11, in the case where f depends on an extra variable $\mathbf{z}_m$, and $n = 1$, $t_i^n = t$ and $t_{i-1}^n = s$ (so that $\Delta_n = t - s$).

Lemma 3.A.13. *Suppose that Assumption 3.2.5 holds, and let $m \in \mathbb{N}_0$. Define $\mathbf{z}_m = (z_1, \ldots, z_m)$, with the convention $\mathbf{z}_0 = ()$, and assume that*

(i) $f(t, y, x, \mathbf{z}_m; \theta)$ is differentiable with respect to θ on Θ for $t \in (0, \Delta_0)_{\varepsilon_0}$, $x, y \in \mathcal{X}$ and $\tilde{\nu}$-almost all $\mathbf{z}_m \in \mathbb{R}^m$.

(ii) for all compact, convex subsets $K \subseteq \Theta$,

$$\sup_{t \in (0, \Delta_0)_{\varepsilon_0}, \theta \in K} \|\partial_\theta f(t, y, x, \mathbf{z}_m; \theta)\|^2 \leq C_{K,m} \left(1 + |x|^{C_{K,m}} + |y|^{C_{K,m}} \right) \prod_{j=1}^{m} \left(1 + |z_j|^{C_{K,m}} \right)$$

for $x, y \in \mathcal{X}$ and $\tilde{\nu}$-almost all $\mathbf{z}_m \in \mathbb{R}^m$.

Let $Df(\,\cdot\,; \theta, \theta') = f(\,\cdot\,; \theta) - f(\,\cdot\,; \theta')$. Then, for $p = 2^q$ with $q \in \mathbb{N}$, there exist constants $C_{K,m,p} > 0$ such that

$$\mathbb{E}_{\theta_0} \left(Df(u - s, X_u, X_s, \mathbf{z}_m; \theta, \theta')^p \right) \leq C_{K,m,p} \, \|\theta - \theta'\|^p \prod_{j=1}^{m} \left(1 + |z_j|^{C_{K,m,p}} \right)$$

for $0 \leq s < u \leq s + \Delta_0$, all $\theta, \theta' \in K$ and $\tilde{\nu}$-almost all $\mathbf{z}_m \in \mathbb{R}^m$. ◇

Lemma 3.A.13 follows by application of the mean value theorem and the Cauchy-Schwarz inequality.

Definition 3.A.14. Define, for $m \in \mathbb{N}$, $\mathbf{z}_m = (z_1, \ldots, z_m)^\star \in \mathbb{R}^m$ and the functions $\tau_m : \mathcal{X} \times \mathbb{R}^m \times \Theta \to \mathcal{X}$ by

$$\tau_m(y, \mathbf{z}_m; \theta) = \tau_{m-1}(y + c(y, z_m; \theta), \mathbf{z}_{m-1}; \theta)$$

where $\mathbf{z}_0 = ()$ and $\tau_0(y, \mathbf{z}_0; \theta) = y$, so that, e.g.

$$\begin{aligned} \tau_1(y, \mathbf{z}_1; \theta) &= y + c(y, z_1; \theta) \\ \tau_2(y, \mathbf{z}_2; \theta) &= y + c(y, z_2; \theta) + c(y + c(y, z_2; \theta), z_1; \theta) . \end{aligned}$$

$\diamond$

Definition 3.A.14 is a slight generalisation of Definition 3.4.9, for use in the following.

Remark 3.A.15. Suppose that Assumption 3.2.5 holds, and let $\tau_m(y, \mathbf{z}_m; \theta_0)$ be as defined by Definition 3.A.14.

(i) It may be seen by induction that for any $m \in \mathbb{N}$, there exist constants $C_m > 0$ such that

$$|\tau_m(y, \mathbf{z}_m; \theta_0)| \leq C_m(1 + |y|) \prod_{j=1}^{m} (1 + |z_j|^{C_m}) .$$

So, for $f(t, y, x; \theta) \in C^{\text{pol}}_{0,0,0,0}((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ and $K \subseteq \Theta$ compact and convex, it may be verified that there exist constants $C_{K,m} > 0$, such that

$$|f(t, \tau_m(y, \mathbf{z}_m; \theta_0), x; \theta)| \leq C_{K,m} \left(1 + |x|^{C_{K,m}} + |y|^{C_{K,m}}\right) \prod_{j=1}^{m} (1 + |z_j|^{C_{K,m}})$$

for all $t \in (0, \Delta_0)_{\varepsilon_0}$, $\theta \in K$, $x, y \in \mathcal{X}$ and $\mathbf{z}_m \in \mathbb{R}^m$. That is, for each $m \in \mathbb{N}$, also

$$((t, y, x; \theta) \mapsto f(t, \tau_m(y, \mathbf{z}_m; \theta_0), x; \theta)) \in C^{\text{pol}}_{0,0,0,0}((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta) .$$

(ii) Suppose that $f(t, y, x; \theta) \in C^{\text{pol}}_{1,2,0,0}((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta)$ and $z \in \mathbb{R}$. Then,

$$\begin{aligned} &\partial_y \left(f(t, y + c(y, z; \theta_0), x; \theta) \right) \\ &\quad = \partial_y f(t, y + c(y, z; \theta_0), x; \theta)(1 + \partial_y c(y, z; \theta_0)) \\ &\partial_y^2 \left(f(t, y + c(y, z; \theta_0), x; \theta) \right) \\ &\quad = \partial_y^2 f(t, y + c(y, z; \theta_0), x; \theta)(1 + \partial_y c(y, z; \theta_0))^2 \\ &\qquad + \partial_y f(t, y + c(y, z; \theta_0), x; \theta) \partial_y^2 c(y, z; \theta_0) , \end{aligned}$$

and (i) may be used to conclude that for fixed $z_1 \in \mathbb{R}$,

$$(t, y, x; \theta) \mapsto f(t, \tau_1(y, z_1; \theta_0), x; \theta) \in C^{\text{pol}}_{1,2,0,0} \left((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta \right) .$$

Using the argument iteratively, it is seen that for fixed $\mathbf{z}_m \in \mathbb{R}^m$, $m \in \mathbb{N}$,

$$(t, y, x; \theta) \mapsto f(t, \tau_m(y, \mathbf{z}_m; \theta_0), x; \theta) \in C^{\text{pol}}_{1,2,0,0} \left((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta \right) .$$

○

Lemma 3.A.16. *Suppose that Assumption 3.2.5 holds, and that*

$$f(t,y,x;\theta) \in C^{pol}_{1,2,0,1}\left((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta\right)$$

with $f(0,x,x;\theta) = 0$ *for all* $x \in \mathcal{X}$ *and* $\theta \in \Theta$. *Let*

$$\zeta_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^n f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta).$$

Then, for $p > d$ *of the form* $p = 2^q$ *for some* $q \in \mathbb{N}$, *the following holds: For each compact, convex set* $K \subseteq \Theta$ *there exists* $C_{K,p} > 0$ *such that*

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^p\right) \leq C_{K,p}\|\theta - \theta'\|^p$$

for all $\theta, \theta' \in K$ *and* $n \in \mathbb{N}$. ◇

Proof of Lemma 3.A.16. Let $K \subseteq \Theta$ compact and convex be given. Choose $p > d$ of the form $p = 2^q$ for some $q \in \mathbb{N}$, and note that

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^p\right) = (n\Delta_n)^{-p}\, \mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^p\right). \tag{3.A.39}$$

By Lemma 3.A.11, there exist constants $C_p > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^p\right) \\
&\quad \leq (n\Delta_n)^{p-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(Df_1(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}; \theta, \theta')^p\right) du \\
&\qquad + (n\Delta_n)^{p/2-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(Df_2(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}; \theta, \theta')^p\right) du \\
&\qquad + \sum_{l=1}^q (n\Delta_n)^{2^{q-l}-1} C_p \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(Df_3(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}, z; \theta, \theta')^p\right) \nu_{\theta_0}(dz)\, du
\end{aligned} \tag{3.A.40}$$

for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$, where f_1, f_2 and f_3 are given by (3.A.32), and $\partial_\theta f_1(t,y,x;\theta)$, $\partial_\theta f_2(t,y,x;\theta)$ and $\partial_\theta f_3(t,y,x,z;\theta)$, well-defined by assumption, satisfy that

$$\begin{aligned}
&\|\partial_\theta f_1(t,y,x;\theta)\|^2 + \|\partial_\theta f_2(t,y,x;\theta)\|^2 + \|\partial_\theta f_3(t,y,x,z;\theta)\|^2 \\
&\quad \leq C_K\left(1 + |x|^{C_K} + |y|^{C_K}\right)\left(1 + |z|^{C_K}\right)
\end{aligned}$$

for all $t \in (0,\Delta_0)_{\varepsilon_0}$ and $\theta \in K$ (see Lemma 3.A.1 and Remark 3.A.15). Then, by Lemma 3.A.13, there exist constants $C_{K,p} > 0$ such that

$$\int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(Df_j(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}; \theta, \theta')^p\right) du \leq C_{K,p}\Delta_n\, \|\theta - \theta'\|^p \tag{3.A.41}$$

for $j = 1, 2$ and

$$\begin{aligned}
&\int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E}_{\theta_0} \left(Df_3(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}, z; \theta, \theta')^p \right) \nu_{\theta_0}(dz)\, du \\
&\quad \le \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} C_{K,p} \|\theta - \theta'\|^p \left(1 + |z|^{C_{K,p}} \right) \nu_{\theta_0}(dz)\, du \\
&\quad \le C_{K,p} \Delta_n \|\theta - \theta'\|^p \,. \qquad (3.A.42)
\end{aligned}$$

Inserting (3.A.41) and (3.A.42) into (3.A.40), yields the existence of $C_{p,K} > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0} \left(\left| \sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta') \right|^p \right) \\
&\quad \le C_{p,K} \left((n\Delta_n)^p + (n\Delta_n)^{p/2} + \sum_{l=1}^q (n\Delta_n)^{2^{q-l}} \right) \|\theta - \theta'\|^p \\
&\quad \le C_{p,K} (n\Delta_n)^p \|\theta - \theta'\|^p \,, \qquad (3.A.43)
\end{aligned}$$

since $n\Delta_n \to \infty$ as $n \to \infty$. Inserting (3.A.43) into (3.A.39) proves Lemma 3.A.16. □

Definition 3.A.17. Suppose that Assumption 3.2.5 holds, and that for some $m \in \mathbb{N}_0$,

$$(t, y, x; \theta) \mapsto f(t, y, x, \mathbf{z}_m; \theta)$$

satisfies Assumption 3.A.10 for $\tilde{\nu}$-almost all $\mathbf{z}_m = (z_1, \dots, z_m) \in \mathbb{R}^m$, with the convention $\mathbf{z}_0 = ()$. Define $\mathcal{A}_1$, $\mathcal{A}_2$ and $\mathcal{A}_3$ by their actions on f, respectively, (f, z), $z \in \mathbb{R}$, as the functions

$$\begin{aligned}
\mathcal{A}_1 f:& \quad (t, y, x, \mathbf{z}_m; \theta) \mapsto \partial_t f(t, y, x, \mathbf{z}_m; \theta) + \mathcal{L}_{\theta_0} f(t, y, x, \mathbf{z}_m; \theta) \\
\mathcal{A}_2 f:& \quad (t, y, x, \mathbf{z}_m; \theta) \mapsto b(y; \theta_0) \partial_y f(t, y, x, \mathbf{z}_m; \theta) \\
\mathcal{A}_3(f, z):& \quad (t, y, x, \mathbf{z}_m, z; \theta) \mapsto f(t, \tau_1(y, z; \theta_0), x, \mathbf{z}_m; \theta) - f(t, y, x, \mathbf{z}_m; \theta) \,.
\end{aligned}$$

When well-defined, let

$$\begin{aligned}
\mathcal{A}_j^k f &= \mathcal{A}_j(\mathcal{A}_j^{k-1} f) \,, \\
\mathcal{A}_3^k(f, \mathbf{w}_k) &= \mathcal{A}_3(\mathcal{A}_3^{k-1}(f, \mathbf{w}_{k-1}), w_k)
\end{aligned}$$

for $j = 1, 2$, with $\mathcal{A}_j^0 f = \mathcal{A}_3^0(f, \mathbf{w}_0) = f$ and $\mathbf{w}_k = (w_1, \dots, w_k) \in \mathbb{R}^k$, $k \in \mathbb{N}$. ◇

Remark 3.A.18. Definition 3.A.17 is used to define a number of functions used in the following. Whenever well-defined for some function $f(t, y, x; \theta)$, the following notation is used, for $j, k = 1, 2$.

$$
\begin{array}{lll lll}
f_j & = & \mathcal{A}_j f & f_3 & = & \mathcal{A}_3(f, z_1) \\
f_{jk} & = & \mathcal{A}_k \mathcal{A}_j f & f_{3k} & = & \mathcal{A}_k \mathcal{A}_3(f, z_1) \\
f_{j2k} & = & \mathcal{A}_k \mathcal{A}_2 \mathcal{A}_j f & f_{32k} & = & \mathcal{A}_k \mathcal{A}_2 \mathcal{A}_3(f, z_1) \\
f_{j23} & = & \mathcal{A}_3(\mathcal{A}_2 \mathcal{A}_j f, z_1) & f_{323} & = & \mathcal{A}_3(\mathcal{A}_2 \mathcal{A}_3(f, z_1), z_2) \\
f_{j23k} & = & \mathcal{A}_k \mathcal{A}_3(\mathcal{A}_2 \mathcal{A}_j f, z_1) & f_{323k} & = & \mathcal{A}_k \mathcal{A}_3(\mathcal{A}_2 \mathcal{A}_3(f, z_1), z_2) \\
f_{j233} & = & \mathcal{A}_3^2(\mathcal{A}_2 \mathcal{A}_j f, \mathbf{z}_2) & f_{3233} & = & \mathcal{A}_3^2(\mathcal{A}_2 \mathcal{A}_3(f, z_1), (z_2, z_3)) \\
f_{j3} & = & \mathcal{A}_3(\mathcal{A}_j f, z_1) & f_{33} & = & \mathcal{A}_3^2(f, \mathbf{z}_2) \\
f_{j3k} & = & \mathcal{A}_k \mathcal{A}_3(\mathcal{A}_j f, z_1) & f_{33k} & = & \mathcal{A}_k \mathcal{A}_3^2(f, \mathbf{z}_2) \\
f_{j32k} & = & \mathcal{A}_k \mathcal{A}_2 \mathcal{A}_3(\mathcal{A}_j f, z_1) & f_{332k} & = & \mathcal{A}_k \mathcal{A}_2 \mathcal{A}_3^2 f, \mathbf{z}_2) \\
f_{j323} & = & \mathcal{A}_3(\mathcal{A}_2 \mathcal{A}_3(\mathcal{A}_j f, z_1), z_2) & f_{3323} & = & \mathcal{A}_3(\mathcal{A}_2 \mathcal{A}_3^2(f, \mathbf{z}_2), z_3) \\
f_{j33} & = & \mathcal{A}_3^2(\mathcal{A}_j f, \mathbf{z}_2) & f_{333} & = & \mathcal{A}_3^3(f, \mathbf{z}_3) \\
f_{j33k} & = & \mathcal{A}_k \mathcal{A}_3^2(\mathcal{A}_j f, \mathbf{z}_2) & f_{333k} & = & \mathcal{A}_k \mathcal{A}_3^3(f, \mathbf{z}_3) \\
f_{j333} & = & \mathcal{A}_3^3(\mathcal{A}_j f, \mathbf{z}_3) & f_{3333} & = & \mathcal{A}_3^4(f, \mathbf{z}_4) \\
f_{j333k} & = & \mathcal{A}_k \mathcal{A}_3^3(\mathcal{A}_j f, \mathbf{z}_3) & f_{3333k} & = & \mathcal{A}_k \mathcal{A}_3^4(f, \mathbf{z}_4) \\
f_{j3333} & = & \mathcal{A}_3^4(\mathcal{A}_j f, \mathbf{z}_4) & f_{33333} & = & \mathcal{A}_3^5(f, \mathbf{z}_5)
\end{array}
$$

For any of these functions, let $m \in \{0, 1, 2, 3, 4, 5\}$ be the number of times $\mathcal{A}_3$ is applied in the function definition. Then the resulting function is a function of $(t, y, x, \mathbf{z}_m; \theta)$. ○

Lemma 3.A.19. *Suppose that Assumption 3.2.5 holds, and that*

$$f(t, y, x; \theta) \in C^{pol}_{2,4,0,1}\left((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta\right)$$

with

$$
\begin{aligned}
f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1, 2 \\
\partial_t f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1 \\
\partial_y f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1 \\
\partial_y^2 f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1\,,
\end{aligned}
$$

and $\tau_k(x, \mathbf{z}_k; \theta_0)$ defined by Definition 3.A.14. Let

$$\zeta_n(\theta) = \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^{n} f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)\,.$$

Then, for any compact, convex set $K \subseteq \Theta$, there exists a constant $C_K > 0$ such that

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^4\right) \leq C_K \|\theta - \theta'\|^4$$

for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$. ◇

Proof of Lemma 3.A.19. In the following, in order to save space, write

$$Dh(\,\cdot\,; \theta, \theta') = h(\,\cdot\,; \theta) - h(\,\cdot\,; \theta')$$

for any function $h(\,\cdot\,;\theta)$, and, for any function of the form $h(t,y,x,\mathbf{z}_m;\theta)$, put

$$\mathbf{h}(u,t_{i-1}^n,\mathbf{z}_m;\theta)=h(u-t_{i-1}^n,X_u,X_{t_{i-1}^n},\mathbf{z}_m;\theta)\,.$$

For $j=1,2,3$, the functions f_j, f_{j1}, f_{j2}, f_{j3}, f_{j31}, f_{j32}, f_{j33} used in the following are defined in Remark 3.A.18. Under the assumptions of this lemma, it may be verified that f, f_j, and f_{j3} satisfy Assumption 3.A.10 (see also Remark 3.A.15.(ii)), and f_{j1}, f_{j2}, f_{j31}, f_{j32}, f_{j33} satisfy the assumptions of Lemma 3.A.13.

Write

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta)-\zeta_n(\theta')|^4\right)=(n\Delta_n)^{-4}\Delta_n^{-2}\,\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta,\theta')\right|^4\right). \tag{3.A.44}$$

By Lemma 3.A.11, there exist constants $C>0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta,\theta')\right|^4\right)\\
&\quad\leq (n\Delta_n)^3C\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\mathbb{E}_{\theta_0}\left(D\mathbf{f_1}(u_1,t_{i-1}^n;\theta,\theta')^4\right)du_1\\
&\qquad+n\Delta_nC\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\mathbb{E}_{\theta_0}\left(D\mathbf{f_2}(u_1,t_{i-1}^n;\theta,\theta')^4\right)du_1\\
&\qquad+(1+n\Delta_n)\,C\sum_{i=1}^n\int_{t_{i-1}^n}^{t_i^n}\int_{\mathbb{R}}\mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1,t_{i-1}^n,z_1;\theta,\theta')^4\right)\nu_{\theta_0}(dz_1)\,du_1
\end{aligned}\tag{3.A.45}$$

for all $\theta,\theta'\in\Theta$ and $n\in\mathbb{N}$. Furthermore, applying Lemma 3.A.12 twice consecutively, there exist constants $C>0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_j}(u_1,t_{i-1}^n;\theta,\theta')^4\right)\\
&\quad\leq C(u_1-t_{i-1}^n)^3\int_{t_{i-1}^n}^{u_1}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j1}}(u_2,t_{i-1}^n;\theta,\theta')^4\right)du_2\\
&\qquad+C(u_1-t_{i-1}^n)\int_{t_{i-1}^n}^{u_1}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j2}}(u_2,t_{i-1}^n;\theta,\theta')^4\right)du_2\\
&\qquad+C\left(1+u_1-t_{i-1}^n\right)\\
&\qquad\quad\times\left(\int_{t_{i-1}^n}^{u_1}\int_{\mathbb{R}}(u_2-t_{i-1}^n)^3\int_{t_{i-1}^n}^{u_2}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j31}}(u_3,t_{i-1}^n,z_1;\theta,\theta')^4\right)du_3\,\nu_{\theta_0}(dz_1)\,du_2\right.\\
&\qquad\qquad+\int_{t_{i-1}^n}^{u_1}\int_{\mathbb{R}}(u_2-t_{i-1}^n)\int_{t_{i-1}^n}^{u_2}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j32}}(u_3,t_{i-1}^n,z_1;\theta,\theta')^4\right)du_3\,\nu_{\theta_0}(dz_1)\,du_2\\
&\qquad\qquad+\int_{t_{i-1}^n}^{u_1}\int_{\mathbb{R}}(1+u_2-t_{i-1}^n)\\
&\qquad\qquad\quad\left.\times\int_{t_{i-1}^n}^{u_2}\int_{\mathbb{R}}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j33}}(u_3,t_{i-1}^n,\mathbf{z}_2;\theta,\theta')^4\right)\nu_{\theta_0}(dz_2)\,du_3\;\nu_{\theta_0}(dz_1)\,du_2\right)
\end{aligned}\tag{3.A.46}$$

for $j = 1, 2$ and

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^4\right) \\
&\quad \leq C(u_1 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_1} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{31}}(u_2, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{32}}(u_2, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_2 \\
&\qquad + C\left(1 + u_1 - t_{i-1}^n\right) \\
&\qquad\quad \times \Bigg(\int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{331}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_3\, \nu_{\theta_0}(dz_2)\, du_2 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{332}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_3\, \nu_{\theta_0}(z_2)\, du_2 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (1 + u_2 - t_{i-1}^n) \\
&\qquad\qquad\quad \times \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) \nu_{\theta_0}(dz_3)\, du_3\;\, \nu_{\theta_0}(dz_2)\, du_2 \Bigg).
\end{aligned} \tag{3.A.47}$$

Let $K \subseteq \Theta$ compact and convex be given. By Lemma 3.A.13, there exist constants $C_K > 0$ such that for $i = 1, \ldots, n$, and

$$\begin{aligned}
j_0 &\in \{11, 12, 21, 22\} \\
j_1 &\in \{31, 32, 131, 132, 231, 232\} \\
j_2 &\in \{133, 233, 331, 332\}
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_0}}(u_3, t_{i-1}^n; \theta, \theta')^4\right) &\leq C_K \|\theta - \theta'\|^4 \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_1}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) &\leq C_K \|\theta - \theta'\|^4 \left(1 + |z_1|^{C_K}\right) \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_2}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) &\leq C_K \|\theta - \theta'\|^4 \prod_{k=1}^{2} \left(1 + |z_k|^{C_K}\right) \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) &\leq C_K \|\theta - \theta'\|^4 \prod_{k=1}^{3} \left(1 + |z_k|^{C_K}\right).
\end{aligned} \tag{3.A.48}$$

Inserting (3.A.48) into (3.A.46) and (3.A.47),

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(D\mathbf{f_j}(u_1, t_{i-1}^n; \theta, \theta')^4\right) &\leq C_K (u_1 - t_{i-1}^n)^2 \|\theta - \theta'\|^4 \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^4\right) &\leq C_K (u_1 - t_{i-1}^n)^2 \left(1 + |z_1|^{C_K}\right) \|\theta - \theta'\|^4
\end{aligned} \tag{3.A.49}$$

for $j = 1, 2$. Inserting (3.A.49) into (3.A.45) yields the existence of $C_K > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^4\right) \\
&\quad \leq C_K \left((n\Delta_n)^4 + (n\Delta_n)^2 + n\Delta_n\right) \Delta_n^2 \|\theta - \theta'\|^4 \\
&\quad \leq C_K (n\Delta_n)^4 \Delta_n^2 \|\theta - \theta'\|^4
\end{aligned} \tag{3.A.50}$$

(recall that $n\Delta_n \to \infty$ as $n \to \infty$). Now, inserting (3.A.50) into (3.A.44), the desired result is obtained. □

Lemma 3.A.20. *Suppose that Assumption 3.2.5 holds, and that*

$$\begin{aligned} f(t,y,x;\theta) &\in C^{pol}_{2,5,0,1}\left((0,\Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta\right) \\ a(y;\theta) &\in C^{pol}_{3,0}(\mathcal{X} \times \Theta) \\ b(y;\theta) &\in C^{pol}_{3,0}(\mathcal{X} \times \Theta) \\ c(y,z;\theta) &\in C^{pol}_{3,0}(\mathcal{X} \times \mathbb{R} \times \Theta) \end{aligned}$$

with

$$\begin{aligned} f(0,\tau_k(x,\mathbf{z}_k;\theta_0),x;\theta) &= 0\,, \quad k = 0,1,2 \\ \partial_t f(0,\tau_k(x,\mathbf{z}_k;\theta_0),x;\theta) &= 0\,, \quad k = 0,1 \\ \partial_y f(0,\tau_k(x,\mathbf{z}_k;\theta_0),x;\theta) &= 0\,, \quad k = 0,1 \\ \partial_y^2 f(0,\tau_k(x,\mathbf{z}_k;\theta_0),x;\theta) &= 0\,, \quad k = 0,1 \\ \partial_y^3 f(0,\tau_k(x,\mathbf{z}_k;\theta_0),x;\theta) &= 0\,, \quad k = 0 \\ \partial_t\partial_y f(0,\tau_k(x,\mathbf{z}_k;\theta_0),x;\theta) &= 0\,, \quad k = 0\,, \end{aligned}$$

and $\tau_k(x,\mathbf{z}_k;\theta_0)$ defined by Definition 3.A.14. Let

$$\zeta_n(\theta) = \frac{1}{n\Delta_n^2}\sum_{i=1}^n f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,.$$

Then, for any compact, convex set $K \subseteq \Theta$, there exists a constant $C_K > 0$ such that

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^2\right) \leq C_K \|\theta - \theta'\|^2$$

for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$. ◇

Proof of Lemma 3.A.20. In the following, in order to save space, write

$$Dh(\,\cdot\,;\theta,\theta') = h(\,\cdot\,;\theta) - h(\,\cdot\,;\theta')$$

for any function $h(\,\cdot\,;\theta)$, and, for any function of the form $h(t,y,x,\mathbf{z}_m;\theta)$, put

$$\mathbf{h}(u,t_{i-1}^n,\mathbf{z}_m;\theta) = h(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}, \mathbf{z}_m;\theta)\,.$$

For $j = 1,2,3$, the functions f_j, f_{j1}, f_{j2}, f_{j21}, f_{j22}, f_{j23}, f_{j3}, f_{j31}, f_{j32}, f_{j33} used in the following are defined in Remark 3.A.18. Under the assumptions of this lemma, it may be verified that f, f_j, f_{j2} and f_{j3}, $j = 1,2,3$, satisfy Assumption 3.A.10 (see also Remark 3.A.15.(ii)), and f_{j1}, f_{j21}, f_{j22}, f_{j23}, f_{j31}, f_{j32}, f_{j33}, $j = 1,2,3$ satisfy the assumptions of Lemma 3.A.13.

Write

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^2\right) = (n\Delta_n)^{-2}\Delta_n^{-2}\,\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta,\theta')\right|^2\right)\,. \tag{3.A.51}$$

By Lemma 3.A.11, there exist constants $C > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^2\right) \\
&\quad \leq n\Delta_n C \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(D\mathbf{f_1}(u_1, t_{i-1}^n; \theta, \theta')^2\right) du_1 \\
&\qquad + C \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(D\mathbf{f_2}(u_1, t_{i-1}^n; \theta, \theta')^2\right) du_1 \\
&\qquad + C \sum_{i=1}^{n} \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^2\right) \nu_{\theta_0}(dz_1)\, du_1
\end{aligned} \tag{3.A.52}$$

for all $\theta, \theta' \in \Theta$. Furthermore, using Lemma 3.A.12 three times, there exist constants $C > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_j}(u_1, t_{i-1}^n; \theta, \theta')^2\right) \\
&\quad \leq C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j1}}(u_2, t_{i-1}^n; \theta, \theta')^2\right) du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j21}}(u_3, t_{i-1}^n; \theta, \theta')^2\right) du_3\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j22}}(u_3, t_{i-1}^n; \theta, \theta')^2\right) du_3\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j23}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^2\right) \nu_{\theta_0}(dz_1)\, du_3\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j31}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^2\right) du_3\, \nu_{\theta_0}(dz_1)\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j32}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^2\right) du_3\, \nu_{\theta_0}(dz_1)\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j33}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^2\right) \nu_{\theta_0}(dz_2)\, du_3\, \nu_{\theta_0}(dz_1)\, du_2
\end{aligned} \tag{3.A.53}$$

for $j = 1, 2$ and

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^2\right) \\
&\quad \le C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{31}}(u_2, t_{i-1}^n, z_1; \theta, \theta')^2\right) du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{321}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^2\right) du_3\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{322}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^2\right) du_3\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{323}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^2\right) \nu_{\theta_0}(dz_2)\, du_3\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{331}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^2\right) du_3\, \nu_{\theta_0}(dz_2)\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{332}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^2\right) du_3\, \nu_{\theta_0}(dz_2)\, du_2 \\
&\qquad + C \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^2\right) \nu_{\theta_0}(dz_3)\, du_3\, \nu_{\theta_0}(dz_2)\, du_2\,.
\end{aligned}
\tag{3.A.54}
$$

Let $K \subseteq \Theta$ compact and convex be given. By Lemma 3.A.13, there exists a constant $C_K > 0$ such that for $i = 1, \ldots, n$ and

$$
\begin{aligned}
&j_0 \in \{11, 21, 121, 122, 221, 222\} \\
&j_1 \in \{31, 123, 131, 132, 223, 231, 232, 321, 322\} \\
&j_2 \in \{133, 233, 323, 331, 332\}\,,
\end{aligned}
$$

it holds that

$$
\begin{aligned}
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_0}}(u_3, t_{i-1}^n; \theta, \theta')^2\right) &\le C_K\, \|\theta - \theta'\|^2 \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_1}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^2\right) &\le C_K\, \|\theta - \theta'\|^2 \left(1 + |z_1|^{C_K}\right) \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_2}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^2\right) &\le C_K \|\theta - \theta'\|^2 \prod_{k=1}^{2} \left(1 + |z_k|^{C_K}\right) \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^2\right) &\le C_K \|\theta - \theta'\|^2 \prod_{k=1}^{3} \left(1 + |z_k|^{C_K}\right)\,.
\end{aligned}
\tag{3.A.55}
$$

Inserting (3.A.55) into (3.A.53) and (3.A.54), we obtain

$$
\begin{aligned}
\mathbb{E}_{\theta_0}\left(D\mathbf{f_j}(u_1, t_{i-1}^n; \theta, \theta')^2\right) &\le C_K (u_1 - t_{i-1}^n)^2\, \|\theta - \theta'\|^2 \\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^2\right) &\le C_K (u_1 - t_{i-1}^n)^2 \left(1 + |z_1|^{C_K}\right) \|\theta - \theta'\|^2
\end{aligned}
\tag{3.A.56}
$$

for $j = 1, 2$, (keeping in mind that $\Delta_n \to 0$ as $n \to \infty$ and $u_1 \le t_i^n$). Inserting (3.A.56) into

(3.A.52), we obtain the existence of $C_K > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^2\right) \\
&\leq C_K\left((n\Delta_n)^2 + n\Delta_n\right)\Delta_n^2\, \|\theta - \theta'\|^2 \\
&\leq C_K (n\Delta_n)^2 \Delta_n^2\, \|\theta - \theta'\|^2
\end{aligned} \tag{3.A.57}$$

(recall that $n\Delta_n \to \infty$ as $n \to \infty$). Now, inserting (3.A.57) into (3.A.51), the desired result is obtained □

Lemma 3.A.21. *Suppose that Assumption 3.2.5 holds, and that*

$$\begin{aligned}
f(t, y, x; \theta) &\in C^{pol}_{2,5,0,1}\left((0, \Delta_0)_{\varepsilon_0} \times \mathcal{X}^2 \times \Theta\right) \\
a(y; \theta) &\in C^{pol}_{3,0}\,(\mathcal{X} \times \Theta) \\
b(y; \theta) &\in C^{pol}_{3,0}\,(\mathcal{X} \times \Theta) \\
c(y, z; \theta) &\in C^{pol}_{3,0}\,(\mathcal{X} \times \mathbb{R} \times \Theta)
\end{aligned}$$

with

$$\begin{aligned}
f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1, 2, 3, 4 \\
\partial_t f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1, 2, 3 \\
\partial_y f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1, 2, 3 \\
\partial_y^2 f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1, 2, 3 \\
\partial_y^3 f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1 \\
\partial_t \partial_y f(0, \tau_k(x, \mathbf{z}_k; \theta_0), x; \theta) &= 0\,, \quad k = 0, 1\,,
\end{aligned}$$

and $\tau_k(x, \mathbf{z}_k; \theta_0)$ defined by Definition 3.A.14. Let

$$\zeta_n(\theta) = \frac{1}{n\Delta_n^2}\sum_{i=1}^{n} f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)\,.$$

Then, for any compact, convex set $K \subseteq \Theta$, there exists a constant $C_K > 0$ such that

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^4\right) \leq C_K \|\theta - \theta'\|^4$$

for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$. ◇

Proof of Lemma 3.A.21. In the following, in order to save space, write

$$Dh(\,\cdot\,; \theta, \theta') = h(\,\cdot\,; \theta) - h(\,\cdot\,; \theta')$$

for any function $h(\,\cdot\,; \theta)$, and, for any function of the form $h(t, y, x, \mathbf{z}_m; \theta)$, put

$$\mathbf{h}(u, t_{i-1}^n, \mathbf{z}_m; \theta) = h(u - t_{i-1}^n, X_u, X_{t_{i-1}^n}, \mathbf{z}_m; \theta)\,.$$

For $j = 1, 2, 3$, the functions

$$f_j \quad f_{j1} \quad f_{j2} \quad f_{j21} \quad f_{j22} \quad f_{j23} \quad f_{j231} \quad f_{j232} \quad f_{j233} \quad f_{j3} \quad f_{j31}$$
$$f_{j32} \quad f_{j321} \quad f_{j322} \quad f_{j323} \quad f_{j33} \quad f_{j331} \quad f_{j332} \quad f_{j333} \quad f_{j3331} \quad f_{j3332} \quad f_{j3333}$$

used in the following are defined in Remark 3.A.18. Under the assumptions of this lemma, it may be verified that

$$f_j \quad f_{j2} \quad f_{j23} \quad f_{j3} \quad f_{j32} \quad f_{j33} \quad f_{j333}$$

satisfy Assumption 3.A.10 (see also Remark 3.A.15.(ii)), and

$$f_{j1} \quad f_{j21} \quad f_{j22} \quad f_{j231} \quad f_{j232} \quad f_{j233} \quad f_{j31} \quad f_{j321}$$
$$f_{j322} \quad f_{j323} \quad f_{j331} \quad f_{j332} \quad f_{j3331} \quad f_{j3332} \quad f_{j3333}$$

satisfy the assumptions of Lemma 3.A.13.

Write

$$\mathbb{E}_{\theta_0}\left(|\zeta_n(\theta) - \zeta_n(\theta')|^4\right) = (n\Delta_n)^{-4}\Delta_n^{-4}\,\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^4\right). \tag{3.A.58}$$

By Lemma 3.A.11, there exist constants $C > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^n Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^4\right) \\
&\quad\le (n\Delta_n)^3 C \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(D\mathbf{f_1}(u_1, t_{i-1}^n; \theta, \theta')^4\right) du_1 \\
&\qquad + Cn\Delta_n \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0}\left(D\mathbf{f_2}(u_1, t_{i-1}^n; \theta, \theta')^4\right) du_1 \\
&\qquad + C(1 + n\Delta_n) \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^4\right) \nu_{\theta_0}(dz_1)\, du_1
\end{aligned} \tag{3.A.59}$$

for all $\theta, \theta' \in \Theta$. Also, by Lemma 3.A.12, there exist constants $C > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_j}(u_1, t_{i-1}^n; \theta, \theta')^4\right) \\
&\quad \leq C(u_1 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_1} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j1}}(u_2, t_{i-1}^n; \theta, \theta')^4\right) du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} (u_2 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j21}}(u_3, t_{i-1}^n; \theta, \theta')^4\right) du_3\, du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j22}}(u_3, t_{i-1}^n; \theta, \theta')^4\right) du_3\, du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \\
&\qquad\quad \times \int_{t_{i-1}^n}^{u_1} (1 + u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j23}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) \nu_{\theta_0}(dz_1)\, du_3\, du_2 \\
&\qquad + C(1 + u_1 - t_{i-1}^n) \\
&\qquad\quad \times \Bigg(\int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j31}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_3\, \nu_{\theta_0}(dz_1)\, du_2 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j32}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_3\, \nu_{\theta_0}(dz_1)\, du_2 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (1 + u_2 - t_{i-1}^n) \\
&\qquad\qquad\quad \times \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j33}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \nu_{\theta_0}(dz_2)\, du_3\, \nu_{\theta_0}(dz_1)\, du_2 \Bigg)
\end{aligned} \tag{3.A.60}$$

for $j = 1, 2$ and

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(Df_3(u_1 - t_{i-1}^n, X_{u_1}, X_{t_{i-1}^n}, z_1; \theta, \theta')^4\right) \\
&\quad \le C(u_1 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_1} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{31}}(u_2, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} (u_2 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{321}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_3\, du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_1} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{322}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_3\, du_2 \\
&\qquad + C(u_1 - t_{i-1}^n) \\
&\qquad\quad \times \int_{t_{i-1}^n}^{u_1} (1 + u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{323}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \nu_{\theta_0}(dz_2)\, du_3\, du_2 \\
&\qquad + C(1 + u_1 - t_{i-1}^n) \\
&\qquad\quad \times \Bigg(\int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{331}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_3\, \nu_{\theta_0}(dz_2)\, du_2 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (u_2 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_2} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{332}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_3\, \nu_{\theta_0}(dz_2)\, du_2 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_1} \int_{\mathbb{R}} (1 + u_2 - t_{i-1}^n) \\
&\qquad\qquad\quad \times \int_{t_{i-1}^n}^{u_2} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) \nu_{\theta_0}(dz_3)\, du_3\, \nu_{\theta_0}(dz_2)\, du_2 \,.
\end{aligned}
\tag{3.A.61}
$$

Furthermore,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j23}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) \\
&\quad \le C(u_3 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j231}}(u_4, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_4 \\
&\qquad + C(u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j232}}(u_4, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_4 \\
&\qquad + C(1 + u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j233}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \nu_{\theta_0}(dz_2)\, du_4 \,,
\end{aligned}
\tag{3.A.62}
$$

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j32}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) \\
&\quad \le C(u_3 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j321}}(u_4, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_4 \\
&\qquad + C(u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j322}}(u_4, t_{i-1}^n, z_1; \theta, \theta')^4\right) du_4 \\
&\qquad + C(1 + u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f_{j323}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \nu_{\theta_0}(dz_2)\, du_4 \,,
\end{aligned}
\tag{3.A.63}
$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{j33}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \\
&\quad \leq C(u_3 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{j331}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_4 \\
&\qquad + C(u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{j332}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_4 \\
&\qquad + C(1 + u_3 - t_{i-1}^n) \\
&\qquad\quad \times \Bigg(\int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} (u_4 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_4} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{j3331}}(u_5, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) du_5\, \nu_{\theta_0}(dz_3)\, du_4 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} (u_4 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_4} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{j3332}}(u_5, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) du_5\, \nu_{\theta_0}(dz_3)\, du_4 \\
&\qquad\qquad + \int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} (1 + u_4 - t_{i-1}^n) \\
&\qquad\qquad\quad \times \int_{t_{i-1}^n}^{u_4} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{j3333}}(u_5, t_{i-1}^n, \mathbf{z}_4; \theta, \theta')^4\right) \nu_{\theta_0}(dz_4)\, du_5\, \nu_{\theta_0}(dz_3)\, du_4 \Bigg),
\end{aligned} \tag{3.A.64}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{323}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \\
&\quad \leq C(u_3 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{3231}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_4 \\
&\qquad + C(u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{3232}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_4 \\
&\qquad + C(1 + u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{3233}}(u_4, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) \nu_{\theta_0}(dz_3)\, du_4\,,
\end{aligned} \tag{3.A.65}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{332}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \\
&\quad \leq C(u_3 - t_{i-1}^n)^3 \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{3321}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_4 \\
&\qquad + C(u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{3322}}(u_4, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) du_4 \\
&\qquad + C(1 + u_3 - t_{i-1}^n) \int_{t_{i-1}^n}^{u_3} \int_{\mathbb{R}} \mathbb{E}_{\theta_0}\left(D\mathbf{f}_{\mathbf{3323}}(u_4, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) \nu_{\theta_0}(dz_3)\, du_4\,,
\end{aligned} \tag{3.A.66}$$

and

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3,t_{i-1}^n,\mathbf{z}_3;\theta,\theta')^4\right)\\
&\quad\le C(u_3-t_{i-1}^n)^3\int_{t_{i-1}^n}^{u_3}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{3331}}(u_4,t_{i-1}^n,\mathbf{z}_3;\theta,\theta')^4\right)du_4\\
&\qquad+C(u_3-t_{i-1}^n)\int_{t_{i-1}^n}^{u_3}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{3332}}(u_4,t_{i-1}^n,\mathbf{z}_3;\theta,\theta')^4\right)du_4\\
&\qquad+C(1+u_3-t_{i-1}^n)\\
&\qquad\quad\times\Bigg(\int_{t_{i-1}^n}^{u_3}\int_{\mathbb{R}}(u_4-t_{i-1}^n)^3\int_{t_{i-1}^n}^{u_4}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{33331}}(u_5,t_{i-1}^n,\mathbf{z}_4;\theta,\theta')^4\right)du_5\,\nu_{\theta_0}(dz_4)\,du_4\\
&\qquad\qquad+\int_{t_{i-1}^n}^{u_3}\int_{\mathbb{R}}(u_4-t_{i-1}^n)\int_{t_{i-1}^n}^{u_4}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{33332}}(u_5,t_{i-1}^n,\mathbf{z}_4;\theta,\theta')^4\right)du_5\,\nu_{\theta_0}(dz_4)\,du_4\\
&\qquad\qquad+\int_{t_{i-1}^n}^{u_3}\int_{\mathbb{R}}(1+u_4-t_{i-1}^n)\\
&\qquad\qquad\quad\times\int_{t_{i-1}^n}^{u_4}\int_{\mathbb{R}}\mathbb{E}_{\theta_0}\left(D\mathbf{f_{33333}}(u_5,t_{i-1}^n,\mathbf{z}_5;\theta,\theta')^4\right)\nu_{\theta_0}(dz_5)\,du_5\,\nu_{\theta_0}(dz_4)\,du_4\Bigg)\,.
\end{aligned}\tag{3.A.67}$$

Let $K\subseteq\Theta$ compact and convex be given. By Lemma 3.A.13, there exist constants $C_K>0$ such that for $i=1,\dots,n$, and

$$\begin{aligned}
&j_0\in\{11,21,121,122,221,222\}\\
&j_1\in\{31,131,231,321,322,1231,1232,1321,1322,2231,2232,2321,2322\}\\
&j_2\in\{331,1233,1323,1331,1332,2233,2323,2331,2332,3231,3232,3321,3322\}\\
&j_3\in\{3233,3323,3331,3332,13331,13332,23331,23332\}\\
&j_4\in\{13333,23333,33331,33332\}
\end{aligned}$$

it holds that

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_0}}(u_3,t_{i-1}^n;\theta,\theta')^4\right)&\le C_K\,\|\theta-\theta'\|^4\\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_1}}(u_3,t_{i-1}^n;\theta,\theta')^4\right)&\le C_K\,\|\theta-\theta'\|^4\left(1+|z_1|^{C_K}\right)\\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_2}}(u_3,t_{i-1}^n,\mathbf{z}_2;\theta,\theta')^4\right)&\le C_K\,\|\theta-\theta'\|^4\prod_{k=1}^{2}\left(1+|z_k|^{C_K}\right)\\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_3}}(u_3,t_{i-1}^n,\mathbf{z}_3;\theta,\theta')^4\right)&\le C_K\,\|\theta-\theta'\|^4\prod_{k=1}^{3}\left(1+|z_k|^{C_K}\right)\\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j_4}}(u_3,t_{i-1}^n,\mathbf{z}_4;\theta,\theta')^4\right)&\le C_K\,\|\theta-\theta'\|^4\prod_{k=1}^{4}\left(1+|z_k|^{C_K}\right)\\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_{33333}}(u_3,t_{i-1}^n,\mathbf{z}_5;\theta,\theta')^4\right)&\le C_K\,\|\theta-\theta'\|^4\prod_{k=1}^{5}\left(1+|z_k|^{C_K}\right)\,.
\end{aligned}\tag{3.A.68}$$

Inserting (3.A.68) into (3.A.62), (3.A.63), (3.A.64), (3.A.65), (3.A.66), (3.A.67), it follows that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j23}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) \le C_K \,\|\theta-\theta'\|^4 (u_3 - t_{i-1}^n)\left(1+|z_1|^{C_K}\right)\\
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j32}}(u_3, t_{i-1}^n, z_1; \theta, \theta')^4\right) \le C_K \,\|\theta-\theta'\|^4 (u_3 - t_{i-1}^n)\left(1+|z_1|^{C_K}\right)\\
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{j33}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \le C_K \,\|\theta-\theta'\|^4 (u_3 - t_{i-1}^n)^2 \prod_{k=1}^{2}\left(1+|z_k|^{C_K}\right)\\
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{323}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \le C_K \,\|\theta-\theta'\|^4 (u_3 - t_{i-1}^n) \prod_{k=1}^{2}\left(1+|z_k|^{C_K}\right)\\
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{332}}(u_3, t_{i-1}^n, \mathbf{z}_2; \theta, \theta')^4\right) \le C_K \,\|\theta-\theta'\|^4 (u_3 - t_{i-1}^n) \prod_{k=1}^{2}\left(1+|z_k|^{C_K}\right)\\
&\mathbb{E}_{\theta_0}\left(D\mathbf{f_{333}}(u_3, t_{i-1}^n, \mathbf{z}_3; \theta, \theta')^4\right) \le C_K \,\|\theta-\theta'\|^4 (u_4 - t_{i-1}^n)^2 \prod_{k=1}^{3}\left(1+|z_k|^{C_K}\right)
\end{aligned} \tag{3.A.69}$$

Inserting the expressions from (3.A.69) into (3.A.60) and (3.A.61) (still using (3.A.68) for the remaining terms), it follows that for $j = 1, 2$,

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left(D\mathbf{f_j}(u_1, t_{i-1}^n; \theta, \theta')^4\right) &\le C_K \,\|\theta-\theta'\|^4 (u_1 - t_{i-1}^n)^4\\
\mathbb{E}_{\theta_0}\left(D\mathbf{f_3}(u_1, t_{i-1}^n, z_1; \theta, \theta')^4\right) &\le C_K \,\|\theta-\theta'\|^4 (u_1 - t_{i-1}^n)^4\left(1+|z_1|^{C_K}\right).
\end{aligned} \tag{3.A.70}$$

Finally, inserting (3.A.70) into (3.A.59) yields the existence of $C_K > 0$ such that

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\left|\sum_{i=1}^{n} Df(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta, \theta')\right|^4\right)\\
&\le C_K\left((n\Delta_n)^4 + (n\Delta_n)^2 + n\Delta_n\right)\Delta_n^4\,\|\theta-\theta'\|^4\\
&\le C_K (n\Delta_n)^4 \Delta_n^4\,\|\theta-\theta'\|^4
\end{aligned} \tag{3.A.71}$$

(recall that $n\Delta_n \to \infty$ as $n \to \infty$). Now, inserting (3.A.71) into (3.A.58), the desired result is obtained □

Lemma 3.A.22. *Suppose that Assumption 3.2.5 holds. Let $f(x;\theta) \in C_{0,1}^{pol}(\mathcal{X}\times\Theta)$, and define*

$$F(\theta) = \int_{\mathcal{X}} f(x;\theta)\,\pi_{\theta_0}(dx)\,.$$

For each $m \in \mathbb{N}$ and compact, convex set $K \subseteq \Theta$, there exists a constant $C_{K,m} > 0$ such that for all $\theta, \theta' \in K$,

$$\mathbb{E}_{\theta_0}\left(|F(\theta) - F(\theta')|^m\right) \le C_{K,m}\|\theta-\theta'\|^m\,.$$

◇

Lemma 3.A.22 may be shown by application of Jensen's inequality, the mean value theorem for functions of several variables, and the Cauchy-Schwarz inequality, and by use of the polynomial growth assumptions on the derivative $\partial_\theta f(x;\theta)$.

Lemma 3.A.23. *Let $\lambda \in \Theta$ be given. Suppose that Assumption 3.2.5 holds, and let $m \geq 2$. Then, there exists a constant $C_{\lambda,m} > 0$ (depending also on Δ_0), such that*

$$\mathbb{E}_\lambda\left(|X_{t+\Delta} - X_t|^m \mid X_t\right) \leq C_{\lambda,m}\Delta(1 + |X_t|^m)$$

for $0 \leq t \leq t + \Delta \leq t + \Delta_0$. ◇

Corollary 3.A.24. *Suppose that Assumption 3.2.5 holds. Let $\lambda \in \Theta$ and a compact, convex subset $K \subseteq \Theta$ be given. Suppose that $f(y, x; \theta)$ is of polynomial growth in x and y, uniformly for θ in compact, convex sets. Then, there exists a constant $C_{\lambda,K} > 0$ (also depending on Δ_0), such that for* $0 \leq t \leq t + \Delta \leq t + \Delta_0$,

$$\mathbb{E}_\lambda\left(|f(X_{t+\Delta}, X_t; \theta)| \mid X_t\right) \leq C_{\lambda,K}(1 + |X_t|^{C_{\lambda,K}}).$$

◇

Lemma 3.A.23 and its corollary correspond to Proposition 3.1 of Shimizu and Yoshida (2006), adapted to the current setup. These results are a key element to controlling remainder terms in this paper. Comparing to Kessler (1997, Lemma 6) (see Lemma 2.A.4 and Corollary 2.A.5), the bound in Lemma 3.A.23 is revealed to be weaker for small Δ, than its continuous-diffusion counterpart.

It was seen in the paper of Shimizu and Yoshida (2006) that the proof of Lemma 3.A.23 is very similar to in the continuous case. However, additional measures are needed to control an additional jump-related term. Shimizu and Yoshida (2006, Lemma 4.1) employed a proof technique of Bichteler and Jacod (1983, Lemma (A.14)) to deal with this term. In the following proof of Lemma 3.A.23, we use Lemma 2.1.5 of Jacod and Protter (2012) to the same end.

With reference to Kessler (1997), a very detailed proof of the continuous-diffusion versions of Lemma 3.A.23 and its corollary (albeit with an easily corrected error) exists in Flachs (2011, Lemmas 3.3 & 3.4). Following the lines of the proof given in Flachs (2011), the proof of Lemma 3.A.23 presented below essentially reproduces the corresponding proof of Shimizu and Yoshida (2006, Proposition 3.1).

Proof of Lemma 3.A.23. Let $\mathbf{M}^{(1)} = (M_s^{(1)})_{s\geq 0}$, $\mathbf{M}^{(2)} = (M_s^{(2)})_{s\geq 0}$ and $\mathbf{M}^{(3)} = (M_s^{(3)})_{s\geq 0}$ be given by

$$\begin{aligned}
M_s^{(1)} &= \int_0^s \mathbf{1}_{(t,t+\Delta]}(u)\tilde{a}(X_{u-};\lambda)\,du \\
M_s^{(2)} &= \int_0^s \mathbf{1}_{(t,t+\Delta]}(u)b(X_{u-};\lambda)\,dW_u \\
M_s^{(3)} &= \int_0^s \int_{\mathbb{R}} \mathbf{1}_{(t,t+\Delta]}(u)c(X_{u-}, z;\lambda)\,(N^\lambda - \mu_\lambda)(du, dz).
\end{aligned}$$

Assumption 3.2.5 ensures that $\mathbf{M}^{(2)}$ and $\mathbf{M}^{(3)}$ are $(\mathcal{F}_s)_{s\geq 0}$ martingales by Applebaum (2009, Theorem 4.2.3).

Using (3.2.7), for some $C_m > 0$,

$$|X_{t+\Delta} - X_t|^m \leq C_m(|M^{(1)}_{t+\Delta}|^m + |M^{(2)}_{t+\Delta}|^m + |M^{(3)}_{t+\Delta}|^m)\,. \tag{3.A.72}$$

By Jensen's inequality and the Burkholder-Davis-Gundy inequality, for $m \geq 2$,

$$\begin{aligned}
&\mathbb{E}_\lambda\left(|M^{(1)}_{t+\Delta}|^m \mid X_t\right) \\
&\quad = \mathbb{E}_\lambda\left(\Delta^m \left|\frac{1}{\Delta}\int_t^{t+\Delta} \tilde{a}(X_{u-};\lambda)\,du\right|^m \mid X_t\right) \\
&\quad \leq \mathbb{E}_\lambda\left(\Delta^{m-1}\int_t^{t+\Delta} |\tilde{a}(X_u;\lambda)|^m\,du \mid X_t\right)
\end{aligned} \tag{3.A.73}$$

and

$$\mathbb{E}_\lambda\left(|M^{(2)}_{t+\Delta}|^m \mid X_t\right) \leq C_{\lambda,m}\mathbb{E}_\lambda\left(\Delta^{m/2-1}\int_t^{t+\Delta} |b(X_u;\lambda)|^m\,du \mid X_t\right). \tag{3.A.74}$$

Also by Jensen's inequality, still for $m \geq 2$,

$$\left(\int_t^{t+\Delta}\int_{\mathbb{R}} c^2(X_{u-},z;\lambda)\nu_\lambda(dz)\,du\right)^{m/2} \leq \Delta_0^{m/2-1}\xi(\lambda)^{m/2-1}\int_t^{t+\Delta}\int_{\mathbb{R}} |c(X_{u-},z;\lambda)|^m\,\nu_\lambda(dz)\,du\,, \tag{3.A.75}$$

where it was used that $\xi(\lambda)^{-1}\nu_\lambda$ is a probability measure. Applying Lemma 2.1.5 of Jacod and Protter (2012), which they prove using Hölder's and Burkholder-Davis-Gundy's inequalities, and inserting (3.A.75),

$$\begin{aligned}
&\mathbb{E}_\lambda\left(|M^{(3)}_{t+\Delta}|^m \mid X_t\right) \\
&\quad \leq C_{\lambda,m}\mathbb{E}_\lambda\left(\int_t^{t+\Delta}\int_{\mathbb{R}} |c(X_{u-},z;\lambda)|^m\,\nu_\lambda(dz)\,du + \left(\int_t^{t+\Delta}\int_{\mathbb{R}} c^2(X_{u-},z;\lambda)\,\nu_\lambda(dz)\,du\right)^{m/2} \mid X_t\right) \\
&\quad \leq C_{\lambda,m}\mathbb{E}_\lambda\left(\int_t^{t+\Delta}\int_{\mathbb{R}} |c(X_{u-},z;\lambda)|^m\,\nu_\lambda(dz)\,du \mid X_t\right).
\end{aligned} \tag{3.A.76}$$

Combining (3.A.73), (3.A.74) and (3.A.76) with (3.A.72), and using that $\tilde{a}(y;\lambda)$, $b(y;\lambda)$ and $c(y,z;\lambda)$ are of linear growth in y,

$$\begin{aligned}
&\mathbb{E}_\lambda\left(|X_{t+\Delta} - X_t|^m \mid X_t\right) \\
&\quad \leq C_{\lambda,m}\mathbb{E}_\lambda\left(\int_t^{t+\Delta}\left(\Delta^{m-1}|\tilde{a}(X_u;\lambda)|^m + \Delta^{m/2-1}|b(X_u;\lambda)|^m + \int_{\mathbb{R}} |c(X_{u-},z;\lambda)|^m\nu_\lambda(dz)\right)du \mid X_t\right) \\
&\quad \leq C_{\lambda,m}\mathbb{E}_\lambda\left(\int_t^{t+\Delta}(1 + |X_t|^m)\,du + \int_t^{t+\Delta}|X_u - X_t|^m\,du \mid X_t\right) \\
&\quad = C_{\lambda,m}\left(\Delta(1 + |X_t|^m) + \int_0^\Delta \mathbb{E}_\lambda\left(|X_{t+u} - X_t|^m \mid X_t\right)\,du\right)
\end{aligned}$$

for suitable contants $C_m > 0$ depending on Δ_0. Now, the Bellman-Gronwall inequality yields the desired result

$$\mathbb{E}_\lambda\left(|X_{t+\Delta} - X_t|^m \mid X_t\right)$$

$$\leq C_{\lambda,m}\left(\Delta(1+|X_t|^m)+\int_0^{\Delta} u(1+|X_t|^m)e^{C_{\lambda,m}(\Delta-u)}\,du\right)$$
$$\leq C_{\lambda,m}(1+|X_t|^m)\left(\Delta+e^{C_{\lambda,m}\Delta_0}\int_0^{\Delta} u\,du\right)$$
$$\leq C_{\lambda,m}\Delta(1+|X_t|^m)\,. \qquad \square$$

Note that by Corollary 3.A.24, it holds that under Assumption 3.2.5,

$$\mathbb{E}_\lambda\left(R_\lambda\left(\Delta, X_{t+\Delta}, X_t;\theta\right) \mid X_t\right) = R_\lambda(\Delta, X_t;\theta) \tag{3.A.77}$$

for $0 \leq t \leq t+\Delta \leq t+\Delta_0$ and $\lambda \in \Theta$.

3.A.4 Expansion of Conditional Moments

Lemma 3.A.25. *Suppose that Assumptions 3.2.5 and 3.2.6 hold. Then,*

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad= \Delta_n\left(\mathcal{L}_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - \mathcal{L}_\theta(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)\\
&\qquad+ \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned} \tag{3.A.78}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad= \Delta_n\left(\mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - \partial_\theta\mathcal{L}_\theta(g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)\\
&\qquad+ \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(gg^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad= \Delta_n\mathcal{L}_{\theta_0}(gg^\star(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}_{\theta_0}\left((\partial_\theta g)^2\left(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta\right) \mid X_{t_{i-1}^n}\right) &= \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\\
\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) &= \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\\
\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) &= \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)
\end{aligned} \tag{3.A.79}$$

for all $j_1, j_2, j_3, j_4 = 1, \ldots, d.$ ◇

Proof of Lemma 3.A.25. Note, for use in the following, that

$$\begin{aligned}
&g_{j_1}g_{j_2}(\Delta, y, x;\theta)\\
&\quad= g_{j_1}g_{j_2}(0, y, x;\theta) + \Delta\left(g_{j_1}^{(1)}g_{j_2}(0, y, x;\theta) + g_{j_1}g_{j_2}^{(1)}(y, x;\theta)\right) + \Delta^2 R(\Delta, y, x;\theta)\,,
\end{aligned}$$

and

$$g_{j_1}g_{j_2}g_{j_3}(\Delta, y, x;\theta) = g_{j_1}g_{j_2}g_{j_3}(0, y, x;\theta) + \Delta R(\Delta, y, x;\theta)$$
$$g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta, y, x;\theta) = g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0, y, x;\theta) + \Delta R(\Delta, y, x;\theta)$$

$$\partial_{\theta_k} g_j(\Delta, y, x; \theta) = \partial_{\theta_k} g_j(0, y, x; \theta) + \Delta \partial_{\theta_k} g_j^{(1)}(y, x; \theta) + \Delta^2 R(\Delta, y, x; \theta) .$$

Using Assumption 3.2.6.(ii), (3.A.77), Remark 3.2.10, Lemma 3.2.8 and Lemma 3.2.9 coordinate-wise, write

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \mathbb{E}_{\theta_0}\left(g(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) + \Delta_n \mathbb{E}_{\theta_0}\left(g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n \mathcal{L}_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n \left(g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n}; \theta)\right) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\quad = \Delta_n \left(\mathcal{L}_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - \mathcal{L}_{\theta}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) ,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \mathbb{E}_{\theta_0}\left(\partial_\theta g(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) + \Delta_n \mathbb{E}_{\theta_0}\left(\partial_\theta g^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \partial_\theta g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n \mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n \left(\partial_\theta g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n}; \theta)\right) \\
&\quad = \Delta_n \left(\mathcal{L}_{\theta_0}(\partial_\theta g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - \partial_\theta \mathcal{L}_{\theta}(g(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right) \\
&\qquad + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) ,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(gg^\star(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \mathbb{E}_{\theta_0}\left(gg^\star(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \Delta_n \mathbb{E}_{\theta_0}\left(g^{(1)}g^\star(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) + g(g^{(1)})^\star(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = gg^\star(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n \mathcal{L}_{\theta_0}(gg^\star(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n \left(g^{(1)}g^\star(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + g(g^{(1)})^\star(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta)\right) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\quad = \Delta_n \mathcal{L}_{\theta_0}(gg^\star(0,\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) .
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left((\partial_\theta g)^2 (\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \mathbb{E}_{\theta_0}\left((\partial_\theta g)^2 (0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) + \Delta_n \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = (\partial_\theta g)^2 (0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\quad = \Delta_n R(\Delta_n, X_{t_{i-1}^n}; \theta) ,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) + \Delta_n \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \Delta_n R(\Delta_n, X_{t_{i-1}^n}; \theta) ,
\end{aligned}$$

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) + \Delta_n\mathbb{E}_{\theta}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta)\,. \qquad \square
\end{aligned}$$

Lemma 3.A.26. *Suppose that Assumptions 3.2.5, 3.2.6, 3.4.8, and Condition 3.A.3 hold.*

(i) For $j_1 = 1,\ldots,d$ *and* $j_2 = d_1+1,\ldots,d$,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \Delta_n^2\left(\tfrac{1}{2}\mathcal{L}_{\theta_0}^2\left(g_{j_1}g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + g_{j_1}^{(1)}g_{j_2}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\right)\\
&\qquad + \Delta_n^2\left(\mathcal{L}_{\theta_0}\left(g_{j_1}(0;\theta)g_{j_2}^{(1)}(\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \mathcal{L}_{\theta_0}\left(g_{j_1}^{(1)}(\theta)g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right)\\
&\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\,.
\end{aligned}$$

(ii) In particular, for $j_1, j_2 = d_1+1,\ldots,d$,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \tfrac{1}{2}\Delta_n^2\left(b^4(x;\beta_0) + \tfrac{1}{2}\left(b^2(x;\beta_0) - b^2(x;\beta)\right)^2\right)\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\\
&\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\,.
\end{aligned}$$

$\diamond$

Proof of Lemma 3.A.26. For $j_1, j_2 = 1,\ldots,d$,

$$\begin{aligned}
&g_{j_1}g_{j_2}(\Delta, y, x;\theta)\\
&\quad = g_{j_1}g_{j_2}(0, y, x;\theta) + \Delta\left(g_{j_1}(0,y,x;\theta)g_{j_2}^{(1)}(y,x;\theta) + g_{j_1}^{(1)}(y,x;\theta)g_{j_2}(0,y,x;\theta)\right)\\
&\qquad + \tfrac{1}{2}\Delta^2\left(g_{j_1}(0,y,x;\theta)g_{j_2}^{(2)}(y,x;\theta) + 2g_{j_1}^{(1)}g_{j_2}^{(1)}(y,x;\theta) + g_{j_1}^{(2)}(y,x;\theta)g_{j_2}(0,y,x;\theta)\right)\\
&\qquad + \Delta^3 R(\Delta, y, x;\theta)\,.
\end{aligned}$$

Lemmas 3.2.8, 3.2.9 and 3.A.4 are used to obtain

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n\mathbb{E}_{\theta_0}\left(g_{j_1}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)g_{j_2}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n\mathbb{E}_{\theta_0}\left(g_{j_1}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta)g_{j_2}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \tfrac{1}{2}\Delta_n^2\mathbb{E}_{\theta_0}\left(g_{j_1}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)g_{j_2}^{(2)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2\mathbb{E}_{\theta_0}\left(g_{j_1}^{(1)}g_{j_2}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \tfrac{1}{2}\Delta_n^2\mathbb{E}_{\theta_0}\left(g_{j_1}^{(2)}(X_{t_i^n}, X_{t_{i-1}^n};\theta)g_{j_2}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^3\mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\,,
\end{aligned}$$

and, for $j_1 = 1, \dots, d$ and $j_2 = d_1 + 1, \dots, d$,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = g_{j_1}g_{j_2}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n \mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \tfrac{1}{2}\Delta_n^2 \mathcal{L}_{\theta_0}^2\left(g_{j_1}g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n g_{j_1}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g_{j_2}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n^2 \mathcal{L}_{\theta_0}\left(g_{j_1}(0;\theta) g_{j_2}^{(1)}(\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n g_{j_1}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g_{j_2}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n^2 \mathcal{L}_{\theta_0}\left(g_{j_1}^{(1)}(\theta) g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \tfrac{1}{2}\Delta_n^2 g_{j_1}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g_{j_2}^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n^2 g_{j_1}^{(1)} g_{j_2}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \tfrac{1}{2}\Delta_n^2 g_{j_1}^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) g_{j_2}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\quad = \tfrac{1}{2}\Delta_n^2 \mathcal{L}_{\theta_0}^2\left(g_{j_1}g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n^2 \mathcal{L}_{\theta_0}\left(g_{j_1}(0;\theta) g_{j_2}^{(1)}(\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n^2 \mathcal{L}_{\theta_0}\left(g_{j_1}^{(1)}(\theta) g_{j_2}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n^2 g_{j_1}^{(1)} g_{j_2}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n}; \theta) .
\end{aligned}
$$

Furthermore, inserting from Corollary 3.A.5, for $j_1, j_2 = d_1 + 1, \dots, d$,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \tfrac{1}{2}\Delta_n^2\left(b^4(x;\beta_0) + \tfrac{1}{2}\left(b^2(x;\beta_0) - b^2(x;\beta)\right)^2\right)\partial_y^2 g_{j_1}\partial_y^2 g_{j_2}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&\qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n}; \theta) .
\end{aligned}
$$

□

Lemma 3.A.27. *Suppose that Assumptions 3.2.5, 3.2.6 and 3.4.8, and Condition 3.A.3 hold. Then,*

(i) for $j_1, j_2 = 1, \dots, d$ and $j_3 = d_1 + 1, \dots, d$,

$$\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) .$$

(ii) for $j_1, j_2, j_3 = 1, \dots, d$ and $j_4 = d_1 + 1, \dots, d$,

$$\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) .$$

(iii) for $j_1 = 1, \dots, d$ and $j_2, j_3, j_4 = d_1 + 1, \dots, d$,

$$\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n}; \theta) .$$

◇

Proof of Lemma 3.A.27. Let $j_1, j_2, j_3, j_4 = 1, \ldots, d$ and $J_k = \{1, \ldots, k\}$ for $k = 3, 4$. Under Assumption 3.4.8,

$$\begin{aligned}
&g_{j_1} g_{j_2} g_{j_3}(\Delta, y, x; \theta) \\
&\quad = g_{j_1} g_{j_2} g_{j_3}(0, y, x; \theta) \\
&\qquad + \Delta \sum_{k=1}^{3} \left(g_{j_k}^{(1)}(y, x; \theta) \prod_{m \in J_3 \setminus \{k\}} g_{j_m}(0, y, x; \theta) \right) \\
&\qquad + \Delta^2 R(\Delta, y, x; \theta) ,
\end{aligned} \tag{3.A.80}$$

and

$$\begin{aligned}
&g_{j_1} g_{j_2} g_{j_3} g_{j_4}(\Delta, y, x; \theta) \\
&\quad = g_{j_1} g_{j_2} g_{j_3} g_{j_4}(0, y, x; \theta) \\
&\qquad + \Delta \sum_{k=1}^{4} \left(g_{j_k}^{(1)}(y, x; \theta) \prod_{m \in J_4 \setminus \{k\}} g_{j_m}(0, y, x; \theta) \right) \\
&\qquad + \tfrac{1}{2} \Delta^2 \sum_{k=1}^{4} \left(g_{j_k}^{(2)}(y, x; \theta) \prod_{m \in J_4 \setminus \{k\}} g_{j_m}(0, y, x; \theta) \right) \\
&\qquad + \tfrac{1}{2} \Delta^2 \sum_{k=1}^{4} \sum_{l \in J_4 \setminus \{k\}} \left(g_{j_k}^{(1)} g_{j_l}^{(1)}(y, x; \theta) \prod_{m \in J_4 \setminus \{k,l\}} g_{j_m}(0, y, x; \theta) \right) \\
&\qquad + \Delta^3 R(\Delta, y, x; \theta) .
\end{aligned} \tag{3.A.81}$$

Using (3.A.80) and Lemmas 3.2.8, 3.2.9 and 3.A.6.(ii),

$$\begin{aligned}
&\mathbb{E}_{\theta_0} \left(g_{j_1} g_{j_2} g_{j_3}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n} \right) \\
&\quad = \mathbb{E}_{\theta_0} \left(g_{j_1} g_{j_2} g_{j_3}(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n} \right) \\
&\qquad + \Delta_n \sum_{k=1}^{3} \mathbb{E}_{\theta_0} \left(g_{j_k}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n}; \theta) \prod_{m \in J_3 \setminus \{k\}} g_{j_m}(0, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n} \right) \\
&\qquad + \Delta_n^2 \mathbb{E}_{\theta_0} \left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n} \right) \\
&\quad = g_{j_1} g_{j_2} g_{j_3}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta_n \mathcal{L}_{\theta_0} \left(g_{j_1} g_{j_2} g_{j_3}(0; \theta) \right) (X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n \sum_{k=1}^{3} \left(g_{j_k}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \prod_{m \in J_3 \setminus \{k\}} g_{j_m}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \right) \\
&\qquad + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
&\quad = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta)
\end{aligned}$$

for $j_1, j_2 = 1, \ldots, d$ and $j_3 = d_1 + 1, \ldots, d$, proving Lemma 3.A.27.(i).

Using also (3.A.81) and Lemma 3.A.6, it holds that for $j_1, j_2, j_3 = 1, \ldots, d$ and $j_4 = d_1 + 1, \ldots, d$,

$$\mathbb{E}_{\theta_0} \left(g_{j_1} g_{j_2} g_{j_3} g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n} \right)$$

$$
\begin{aligned}
&= \mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + \Delta_n \sum_{k=1}^{4} \mathbb{E}_{\theta_0}\left(g_{j_k}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \prod_{m\in J_4\setminus\{k\}} g_{j_m}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\quad + \Delta_n^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&= g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n \mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\quad + \Delta_n \sum_{k=1}^{4} \left(g_{j_k}^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) \prod_{m\in J_4\setminus\{k\}} g_{j_m}(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\right) \\
&\quad + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta) \\
&= \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)
\end{aligned}
$$

and Lemma 3.A.27.(ii) follows. Similarly,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \Delta_n \sum_{k=1}^{4} \mathbb{E}_{\theta_0}\left(g_{j_k}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \prod_{m\in J_4\setminus\{k\}} g_{j_m}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \tfrac{1}{2}\Delta_n^2 \sum_{k=1}^{4} \mathbb{E}_{\theta_0}\left(g_{j_k}^{(2)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \prod_{m\in J_4\setminus\{k\}} g_{j_m}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \tfrac{1}{2}\Delta_n^2 \sum_{k=1}^{4} \sum_{l\in J_4\setminus\{k\}} \mathbb{E}_{\theta_0}\left(g_{j_k}^{(1)}g_{j_l}^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \prod_{m\in J_4\setminus\{k,l\}} g_{j_m}(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\qquad + \Delta_n^3 \mathbb{E}_{\theta_0}\left(R(\Delta, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right),
\end{aligned}
$$

so, by Lemmas 3.2.8 and 3.2.9,

$$
\begin{aligned}
&\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&\quad = \Delta_n \mathcal{L}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \tfrac{1}{2}\Delta_n^2 \mathcal{L}_{\theta_0}^2\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n^2 \sum_{k=1}^{4} \mathcal{L}_{\theta_0}\left(g_{j_k}^{(1)}(\theta) \prod_{m\in J_4\setminus\{k\}} g_{j_m}(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) \\
&\qquad + \Delta_n^3 R(\Delta, X_{t_{i-1}^n};\theta),
\end{aligned}
$$

and by Lemma 3.A.6, for $j_1 = 1, \dots, d$ and $j_2, j_3, j_4 = d_1 + 1, \dots, d$,

$$
\mathbb{E}_{\theta_0}\left(g_{j_1}g_{j_2}g_{j_3}g_{j_4}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^3 R(\Delta, X_{t_{i-1}^n};\theta),
$$

proving Lemma 3.A.27.(iii). □

Lemma 3.A.28. *Suppose that Assumptions 3.2.5, 3.2.6, 3.4.8, and Condition 3.A.3 hold, and that $\partial_y^2 \partial_\alpha g_\beta(0, x, x;\theta) = 0$ for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Then*

$$
\mathbb{E}_{\theta_0}\left(\partial_\alpha g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta),
$$

$$\mathbb{E}_{\theta_0}\left(\partial_\alpha g_\beta(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right) = \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n};\theta)\,.$$

◇

Proof of Lemma 3.A.28. For $j = d_1+1,\ldots,d$ and $k = 1,\ldots,d_1$,

$$\begin{aligned}
&\partial_{\theta_k} g_j(\Delta, y, x;\theta)\\
&\quad = \partial_{\theta_k} g_j(0,y,x;\theta) + \Delta\partial_{\theta_k} g_j^{(1)}(y,x;\theta) + \tfrac{1}{2}\Delta^2\partial_{\theta_k} g_j^{(2)}(y,x;\theta) + \Delta^3 R(\Delta,y,x;\theta)
\end{aligned} \tag{3.A.82}$$

and

$$\begin{aligned}
&\partial_{\theta_k} g_j(\Delta, y, x;\theta)^2\\
&\quad = \partial_{\theta_k} g_j(0,y,x;\theta)^2 + 2\Delta\partial_{\theta_k} g_j(0,y,x;\theta)\partial_{\theta_k} g_j^{(1)}(y,x;\theta)\\
&\qquad + \Delta^2\left(\partial_{\theta_k} g_j(0,y,x;\theta)\partial_{\theta_k} g_j^{(2)}(y,x;\theta) + \partial_{\theta_k} g_j^{(1)}(y,x;\theta)^2\right)\\
&\qquad + \Delta^3 R(\Delta,y,x;\theta)\,.
\end{aligned} \tag{3.A.83}$$

Using (3.A.82), (3.A.83) and Lemmas 3.2.8, 3.2.9 and 3.A.7,

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j(0, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) + \Delta_n\mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j^{(1)}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2\mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \partial_{\theta_k} g_j(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n\mathcal{L}_{\theta_0}\left(\partial_{\theta_k} g_j(0;\theta)\right)(X_{t_{i-1}^n}; X_{t_{i-1}^n})\\
&\qquad + \Delta_n\partial_{\theta_k} g_j^{(1)}(\Delta_n, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \partial_{\theta_k} g_j(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + \Delta_n\mathcal{L}_{\theta_0}\left(\partial_{\theta_k} g_j(0;\theta)\right)(X_{t_{i-1}^n}; X_{t_{i-1}^n})\\
&\qquad - \Delta_n\partial_{\theta_k}\mathcal{L}_{\theta}\left(g_j(0;\theta)\right)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\\
&\quad = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)\,,
\end{aligned}$$

and

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right)\\
&\quad = \mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n\mathbb{E}_{\theta_0}\left(2\partial_{\theta_k} g_j(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)\partial_{\theta_k} g_j^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2\mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j(0, X_{t_i^n}, X_{t_{i-1}^n};\theta)\partial_{\theta_k} g_j^{(2)}(X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^2\mathbb{E}_{\theta_0}\left(\partial_{\theta_k} g_j^{(1)}(X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right)\\
&\qquad + \Delta_n^3\mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right)\\
&\quad = \partial_{\theta_k} g_j(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^2 + \Delta_n\mathcal{L}_{\theta_0}\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(X_{t_{i-1}^n}; X_{t_{i-1}^n})\\
&\qquad + \tfrac{1}{2}\Delta_n^2\mathcal{L}_{\theta_0}^2\left((\partial_{\theta_k} g_j)^2(0;\theta)\right)(X_{t_{i-1}^n}; X_{t_{i-1}^n})\\
&\qquad + 2\Delta_n\partial_{\theta_k} g_j(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\partial_{\theta_k} g_j^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\\
&\qquad + 2\Delta_n^2\mathcal{L}_{\theta_0}\left(\partial_{\theta_k} g_j(0;\theta)\partial_{\theta_k} g_j^{(1)}(\theta)\right)(X_{t_{i-1}^n}; X_{t_{i-1}^n})
\end{aligned}$$

$$
\begin{aligned}
&+ \Delta_n^2 \partial_{\theta_k} g_j(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \partial_{\theta_k} g_j^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \\
&+ \Delta_n^2 \partial_{\theta_k} g_j^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta)^2 \\
&+ \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n}; \theta) \\
= {}& \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n}; \theta) . \qquad \square
\end{aligned}
$$

Proof of Lemma 3.2.8

Flachs (2011) gives a detailed proof of Lemma 3.2.8 in the case of ergodic diffusions without jumps, based on the proof in an earlier version of Sørensen (2012, Lemma 1.10), see Flachs (2011, Lemmas 3.7 & 3.8). The proof presented here extends these proofs to cover diffusions with jumps. Although the general Assumption 3.2.5 includes the assumption of ergodicity, this is not actually made use of in the proof below.

Proof of Lemma 3.2.8. Observe first, for use in the following, that if the present assumptions are satisfied for some $k \in \mathbb{N}$, then $\mathcal{L}_\lambda^i f(y, x; \theta) \in C^{\text{pol}}_{2(k+1-i),0,0}(\mathcal{X}^2 \times \Theta)$ for $i = 0, \dots, k+1$ by the help of Lemma 3.A.1. This implies the existence of constants $C_\theta > 0$ such that for $0 \le v \le \Delta$,

$$
\mathbb{E}\left(\int_t^{t+v} |\mathcal{L}_\lambda^i f(X_{s-}, X_t; \theta)| \, ds \right) \le C_\theta \int_t^{t+v} \left(1 + \sup_{u \in [0,\infty)} \mathbb{E}\left(|X_u|^{C_\theta} \right) \right) ds \le C_\theta \Delta_0 , \tag{3.A.84}
$$

allowing for the interchanging of integrals (and conditional expectations).

Furthermore, due to the finite activity of the jumps under consideration, it holds that for fixed $\omega \in \Omega$, $X_t(\omega) \neq X_{t-}(\omega)$ for at most countably many t in any finite interval. Hence X_{s-} may be replaced by X_s in integrals with respect to time, like in the leftmost integral in (3.A.84). In the following, such replacements are often made in integrands which themselves are conditional expectations, by implicitly interchanging the order of the outer integral and the conditional expectation twice

First, the expansion of the conditional expectation in powers of Δ,

$$
\begin{aligned}
\mathbb{E}\left(f(X_{t+\Delta}, X_t; \theta) \mid X_t \right) = {}& \sum_{i=0}^{k} \frac{\Delta^i}{i!} \mathcal{L}^i f(X_t, X_t; \theta) \\
&+ \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E}\left(\mathcal{L}^{k+1} f(X_{(t+u_{k+1})-}, X_t; \theta) \mid X_t \right) du_{k+1} \cdots du_1 ,
\end{aligned}
\tag{3.A.85}
$$

is proven by induction on k, using Itô's formula for stochastic differential equations with jumps, Lemma 3.A.9.

Using the martingale properties of the stochastic integrals, it follows immediately from Lemma 3.A.9 and the previous observations that

$$
\mathbb{E}_\lambda \left(f(X_{t+\Delta}, X_t; \theta) \mid X_t \right) = f(X_t, X_t; \theta) + \mathbb{E}_\lambda \left(\int_t^{t+\Delta} \mathcal{L}_\lambda f(X_s, X_t; \theta) \, ds \mid X_t \right)
$$

$$= f(X_t, X_t; \theta) + \int_0^\Delta \mathbb{E}_\lambda \left(\mathcal{L}_\lambda f(X_{t+s}, X_t; \theta) \mid X_t \right) ds \,,$$

proving (3.A.85) for $k = 0$.

Now, assume that (3.A.85) holds for some $k \in \mathbb{N}_0$, and suppose that the assumptions of the lemma are satisfied for $k + 1$. Then, in particular, $\mathcal{L}_\lambda^{k+1} f(y, x; \theta) \in C^{\text{pol}}_{2,0,0}(\mathcal{X}^2 \times \Theta)$. Using Lemma 3.A.9 again,

$$\begin{aligned}
&\int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E}_\lambda \left(\mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t \right) du_{k+1} \cdots du_1 \\
&\quad = \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathcal{L}_\lambda^{k+1} f(X_t, X_t; \theta) \, du_{k+1} \cdots du_1 \\
&\qquad + \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \int_0^{u_{k+1}} \mathbb{E}_\lambda \left(\mathcal{L}_\lambda^{k+2} f(X_{t+u_{k+2}}, X_t; \theta) \mid X_t \right) du_{k+2} \, du_{k+1} \cdots du_1 \\
&\quad = \frac{\Delta^{k+1}}{(k+1)!} \mathcal{L}_\lambda^{k+1} f(X_t, X_t; \theta) \\
&\qquad + \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_{k+1}} \mathbb{E}_\lambda \left(\mathcal{L}_\lambda^{k+2} f(X_{t+u_{k+2}}, X_t; \theta) \mid X_t \right) du_{k+2} \cdots du_1 \,,
\end{aligned}$$

from which the validity of the expansion follows for $k + 1$, and thus for general $k \in \mathbb{N}_0$ by induction.

It remains to show that for $k \in \mathbb{N}_0$,

$$\int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E} \left(\mathcal{L}^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t \right) du_{k+1} \cdots du_1 = \Delta^{k+1} R(\Delta, X_t; \theta) \,.$$

As seen in the proof of Kessler (1997, Lemma 1) for diffusions without jumps, the remainder term is controlled by an application of Corollary 3.A.24 to $\mathcal{L}_\lambda^{k+1} f$:

Let $k \in \mathbb{N}_0$ be given, so that $\mathcal{L}_\lambda^{k+1} f(y, x; \theta) \in C^{\text{pol}}_{0,0,0}(\mathcal{X}^2 \times \Theta)$, and choose any compact, convex subset $K \subseteq \Theta$. By the corollary, there exist constants $C_K > 0$ such that

$$\left| \mathbb{E}_\lambda \left(\mathcal{L}_\lambda^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t \right) \right| \leq C_K \left(1 + |X_t|^{C_K} \right)$$

for all $\theta \in K$. Then

$$\sup_{\theta \in K} \left| \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E} \left(\mathcal{L}^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t \right) du_{k+1} \cdots du_1 \right| \leq \Delta^{k+1} C_K \left(1 + |X_t|^{C_K} \right) ,$$

i.e.

$$\Delta^{-(k+1)} \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} \mathbb{E} \left(\mathcal{L}^{k+1} f(X_{t+u_{k+1}}, X_t; \theta) \mid X_t \right) du_{k+1} \cdots du_1 = R(\Delta, X_t; \theta) \,,$$

which completes the proof. □

3.A.5 Convergence in Probability

Lemma 3.A.29. *Suppose that Assumption 3.2.5 holds, and that for fixed $\theta \in \Theta$, the functions $x \mapsto f(x;\theta)$ and $x \mapsto \partial_x f(x;\theta)$ are continuous and of polynomial growth in x for $x \in \mathcal{X}$ and . Then*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{t_{i-1}^n};\theta) \xrightarrow{\mathcal{P}} \int_{\mathcal{X}} f(x;\theta)\,\pi_{\theta_0}(dx)\,,$$

point-wise for $\theta \in \Theta$. ◇

Using the ergodicity of $\mathbf{X}$ (Assumption 3.2.5.(v)), Lemma 3.A.23, the Cauchy-Schwarz and Jensen's inequalities, and the assumptions of polynomial growth, Lemma 3.A.29 is proven in the same way as the non-uniform part of Kessler (1997, Lemma 8), see also Masuda (2013, p. 1598). The proof is omitted here.

Remark 3.A.30. For all $\theta \in \Theta$, $|R_\theta(t,x)| \leq C_\theta(1+|x|^{C_\theta})$ for all $t \in (0,\Delta_0)_{\varepsilon_0}$ and $x \in \mathcal{X}$, so, by Lemma 3.A.29, whenever $(\delta_n)_{n\in\mathbb{N}}$ is a sequence of non-negative numbers with $\delta_n \to 0$ as $n \to \infty$,

$$\delta_n \frac{1}{n}\sum_{i=1}^{n} |R_\theta(\Delta_n, X_{t_{i-1}^n})| \leq \delta_n C_\theta \frac{1}{n}\sum_{i=1}^{n}\left(1+|X_{t_{i-1}^n}|^{C_\theta}\right) \xrightarrow{\mathcal{P}} 0\,.$$

In particular, this is also true for $R_\theta(t,x) = R(t,x;\theta)$. ○

Lemma 3.A.31 corresponds to Lemma 9 of Genon-Catalot and Jacod (1993), the proof is omitted here.

Lemma 3.A.31. *For $i = 0,\ldots,n$, $n \in \mathbb{N}$, let $\mathcal{F}_{n,i} = \mathcal{F}_{t_i^n}$, and let $F_{n,i}$ be an $\mathcal{F}_{n,i}$-measurable random variable. If*

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_0}(F_{n,i} \mid \mathcal{F}_{n,i-1}) \xrightarrow{\mathcal{P}} Z \quad \text{and} \quad \sum_{i=1}^{n} \mathbb{E}_{\theta_0}(F_{n,i}^2 \mid \mathcal{F}_{n,i-1}) \xrightarrow{\mathcal{P}} 0\,,$$

for some random variable F, then

$$\sum_{i=1}^{n} F_{n,i} \xrightarrow{\mathcal{P}} F\,.$$

◇

Lemma 3.A.32. *Let $K \subseteq \Theta$ be compact and convex. Suppose that for $n \in \mathbb{N}$, $\mathbf{H}_n = (H_n(\theta))_{\theta\in K}$ is a continuous, real-valued stochastic process, such that*

$$H_n(\theta) \xrightarrow{\mathcal{P}} 0$$

point-wise for $\theta \in K$. Furthermore, assume that there exist constants $p > d$ and $C_{K,p} > 0$ such that for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$,

$$\mathbb{E}_{\theta_0}\left|H_n(\theta) - H_n(\theta')\right|^p \leq C_{p,K}\|\theta-\theta'\|^p\,.$$

Then,

$$\sup_{\theta \in K} |H_n(\theta)| \xrightarrow{\mathcal{P}} 0\,.$$

◇

Lemma 3.A.33. *Suppose that for* $n \in \mathbb{N}$, $\mathbf{H}_n = (H_n(\theta))_{\theta \in K}$ *and* $(H(\theta))_{\theta \in \Theta}$ *are continuous, real-valued stochastic process. If*

$$\sup_{\theta \in K} |H_n(\theta) - H(\theta)| \xrightarrow{\mathcal{P}} 0$$

for all compact, convex sets $K \subseteq \Theta$, *and* $\hat{\theta}_n$ *is a consistent estimator of* θ_0, *then*

$$H_n(\hat{\theta}_n) \xrightarrow{\mathcal{P}} H(\theta_0)\,.$$

◇

Lemmas 3.A.32 and 3.A.33 extend the results of Lemmas 2.A.9 and 2.A.10, and the proofs given in Appendix 2.A easily adapt to the present situation. In particular, Lemma 3.A.32 may be shown using results from Kallenberg (1997, Chapter 14).

3.B Theorems from the Literature

This section summarises some theorems from the literature, which are important to the proofs in Section 3.5. The theorems are presented here without proof, most of them in a greatly simplified form, and tailored specifically to fit the approximate martingale estimating function-setup considered in this paper. Section 3.B.1 contains a version of Corollary 3.1 of Hall and Heyde (1980), while Section 3.B.2 contains selected results from Section 1.10 of Sørensen (2012).

3.B.1 Martingale Central Limit Theorem

This section contains a version of the central limit theorem for martingale differences from Section 3.2 of Hall and Heyde (1980). Recall that we defined $\mathcal{G}_{n,i}$ as the σ-algebra generated by $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_i^n})$. Suppose that for each $n \in \mathbb{N}$, $(M_{n,i})_{1 \le i \le n}$ is a real-valued, zero-mean, square-integrable martingale with respect to $(\mathcal{G}_{n,i})_{1 \le i \le n}$. Let

$$D_{n,i} = M_{n,i} - M_{n,i-1}\,, \quad 1 \le i \le n\,,$$

with $D_{n,0} = 0$ denote the corresponding martingale differences. This collection constitutes a zero-mean, square-integrable martingale array $\{M_{n,i}, \mathcal{G}_{n,i} : 1 \le i \le n, n \in \mathbb{N}\}$ with differences $D_{n,i}$.

Theorem 3.B.1. *(Hall and Heyde, 1980, Corollary 3.1) Suppose that* $\{M_{n,i}, \mathcal{G}_{n,i} : 1 \le i \le n, n \in \mathbb{N}\}$ *is a zero-mean, square-integrable martingale array with differences* $D_{n,i}$. *If*

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left(D_{n,i}^2 \mid \mathcal{G}_{n,i-1}\right) \xrightarrow{\mathcal{P}} C(\theta_0)$$

for some real-valued constant $C(\theta_0)$, and if for all $\varepsilon > 0$,

$$\sum_{i=1}^{n} \mathbb{E}_{\theta_0}\left(D_{n,i}^2 \mathbf{1}(|D_{n,i}| > \varepsilon) \mid \mathcal{G}_{n,i-1}\right) \xrightarrow{\mathcal{P}} 0$$

(the conditional Lindeberg condition holds), then

$$\sum_{i=1}^{n} D_{n,i} \xrightarrow{\mathcal{D}} \mathcal{N}(0, C(\theta_0)) .$$

◇

3.B.2 Asymptotic Results for Estimating Functions

This section briefly summarises Theorems 1.58, 1.59 and 1.60 and some additional comments from Sørensen (2012, Section 1.10), adapted to the setup of the current paper. Proofs of these results are given by Jacod and Sørensen (2012).

In the following, let $G_n(\theta)$ be an approximate martingale estimating function as given in Definition 3.2.3, with associated G_n-estimators defined in Definition 3.2.4.

Theorem 3.B.2. *Sørensen (2012, Theorem 1.58) Suppose that there exist a compact, convex set $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$, and a function $\theta \mapsto B(\theta;\theta_0)$ on K, with values in the set of $d \times d$ matrices, such that*

(i) $G_n(\theta_0) \xrightarrow{\mathcal{P}} 0$.

(ii) The mapping $\theta \mapsto G_n(\theta)$ is continuously differentiable on K for all $n \in \mathbb{N}$ with

$$\sup_{\theta\in K} \|\partial_\theta G_n(\theta) - B(\theta;\theta_0)\| \xrightarrow{\mathcal{P}} 0 .$$

(iii) $B(\theta_0;\theta_0)$ is non-singular.

Then, there exists a consistent G_n-estimator $\hat{\theta}_n$, which is eventually unique in the sense that for any other consistent G_n-estimator $\bar{\theta}_n$, $\mathbb{P}_{\theta_0}(\hat{\theta}_n \neq \bar{\theta}_n) \to 0$ as $n \to \infty$. ◇

By Sørensen (2012, p. 87), under the conditions of Theorem 3.B.2, the mapping $\theta \mapsto B(\theta;\theta_0)$ is continuous on K. Also, there exists a unique, continuously differentiable mapping $\theta \mapsto A(\theta;\theta_0)$ with values in $\mathbb{R}^d$, satisfying that $A(\theta_0;\theta_0) = 0$, $\theta \mapsto \partial_\theta A(\theta;\theta_0) = B(\theta;\theta_0)$ for all $\theta \in K$ and

$$\sup_{\theta\in K} \|G_n(\theta) - A(\theta;\theta_0)\| \xrightarrow{\mathcal{P}} 0 .$$

Theorem 3.B.3. *Sørensen (2012, Theorem 1.59) Suppose that the conditions of Theorem 3.B.2 are satisfied, and that the aforementioned function $A(\theta;\theta_0)$ satisfies that for all $\varepsilon > 0$,*

$$\inf_{K\setminus \bar{B}_\varepsilon(\theta_0)} \|A(\theta;\theta_0)\| > 0 , \tag{3.B.1}$$

where $\bar{B}_\varepsilon(\theta_0)$ *denotes the closed ball with radius* ε *and centre* θ_0. *Then, for any* G_n-*estimator* $\tilde{\theta}_n$, *it holds that for all* $\varepsilon > 0$,

$$\mathbb{P}_{\theta_0}\left(\tilde{\theta}_n \in K \backslash \bar{B}_\varepsilon(\theta_0)\right) \to 0$$

as $n \to \infty$. ◇

Theorem 3.B.4. *Sørensen (2012, Theorem 1.60) Suppose that* $G_n(\theta)$ *satisfies the conditions of Theorem 3.B.2, and let* δ_n *be a sequence of invertible, diagonal* $d \times d$ *matrices, with each entry of* δ_n^{-1} *going to* 0 *as* $n \to \infty$. *Suppose that there exists*

(i) an $\mathbb{R}^d$-*valued random variable* $G(\theta_0)$, *normally distributed with mean zero and positive definite covariance matrix* $J(\theta_0)$, *such that*

$$\delta_n G_n(\theta_0) \xrightarrow{\mathcal{D}} G(\theta_0)\,.$$

(ii) a deterministic function $\theta \mapsto H(\theta; \theta_0)$ *on* K, *with values in the set of* $d \times d$ *matrices and* $H(\theta_0; \theta_0)$ *invertible, such that*

$$\sup_{\theta \in K} \|\delta_n \partial_\theta G_n(\theta) \delta_n^{-1} - H(\theta; \theta_0)\| \xrightarrow{\mathcal{P}} 0\,.$$

Then, for any consistent G_n-*estimator* $\hat{\theta}_n$,

$$\delta_n(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}_d(0, H(\theta_0; \theta_0)^{-1} J(\theta_0)(H(\theta_0; \theta_0)^\star)^{-1})\,.$$

◇

Bibliography

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed form approximation approach. *Econometrica*, 70(1):223–262.

Aït-Sahalia, Y. (2008). Closed form likelihood expansions for multivariate diffusions. *The Annals of Statistics*, 36(2):906–937.

Aldous, D. J. and Eagleson, G. K. (1978). On mixing and stability of limit theorems. *The Annals of Probability*, 6(2):325–331.

Applebaum, D. (2009). *Lévy Processes and Stochastic Calculus*. Number 116 in Cambridge studies in advanced mathematics. Cambridge University Press, 2nd edition.

Bandi, F. M. and Nguyen, T. H. (2003). On the functional estimation of jump-diffusion models. *Journal of Econometrics*, 116(1-2):293–328.

Bandi, F. M. and Phillips, P. C. B. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1):241–283.

Becheri, I. G., Drost, F. C., and Werker, B. J. M. (2014). Asymptotic inference for jump diffusions with state-dependent intensity. Available for download at SSRN: http://ssrn.com/abstract=2292486.

Beskos, A., Papaspiliopoulos, O., and Roberts, G. (2009). Monte Carlo maximum likelihood estimation for discretely observed diffusion processes. *The Annals of Statistics*, 37(1):223–245.

Beskos, A., Papaspiliopoulos, O., Roberts, G. O., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(3):333–382.

Bibbona, E., Lansky, P., and Sirovich, R. (2010). Estimating input parameters from intracellular recordings in the Feller neuronal model. *Physical Review E*, 81(3):031916.

Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, 1(1-2):17–39.

Bichteler, K. and Jacod, J. (1983). Calcul de Malliavin pour les diffusions avec sauts: existence d'une densité dans le cas unidimensionnel. *Séminaire de probabilités de Strasbourg*, 17:132–157.

Clément, E. and Gloter, A. (2015). Local asymptotic mixed normality property for discretely observed stochastic differential equations driven by stable Lévy processes. Stochastic Processes and their Applications (accepted by journal).

Comte, F., Genon-Catalot, V., and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, 13(2):514–543.

Cox, J. C., Ingersoll, Jr., J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics - An International Journal of Probability and Stochastic Processes*, 19(4):263–284.

De Jong, F., Drost, F. C., and Werker, B. J. M. (2001). A jump-diffusion model for exchange rates in a target zone. *Statistica Neerlandica*, 55(3):270–300.

Ditlevsen, S. and Lansky, P. (2006). Estimation of the input parameters in the Feller neuronal model. *Physical Review E*, 73(6):061910.

Dohnal, G. (1987). On estimating the diffusion coefficient. *Journal of Applied Probability*, 24(1):105–114.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–316.

Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. *Scandinavian Journal of Statistics*, 37(2):200–220.

Filipović, D., Mayerhofer, E., and Schneider, P. (2013). Density approximations for multivariate affine jump-diffusion processes. *Journal of Econometrics*, 176(2):93–111.

Flachs, F. (2011). Efficient estimation af parametre i diffusionskoefficienten for en højfrekvensiel diffusion. Master's thesis, University of Copenhagen, Denmark.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 20(4):547–557.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, 30(4):790–804.

Forman, J. L. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35(3):438–465.

Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1):99–116.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29(1):119–151.

Genon-Catalot, V. and Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: Random sampling. *Scandinavian Journal of Statistics*, 21(3):193–221.

Genon-Catalot, V., Laredo, C., and Picard, D. (1992). Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scandinavian Journal of Statistics*, 19(4):317–335.

Giesecke, K. and Schwenkler, G. (2014). Simulated likelihood estimators for discretely observed jump-diffusions. Available at SSRN: http://ssrn.com/abstract=2518160.

Giraudo, M. T. and Sacerdote, L. (1997). Jump-diffusion processes as models for neuronal activity. *Biosystems*, 40(1-2):75–82.

Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699.

Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6):899–912.

Gobet, E. (2002). LAN property for ergodic diffusions with discrete observations. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 38(5):711–737.

Godambe, V. P. and Heyde, C. C. (1987). Quasi-likelihood and optimal estimation. *International Statistical Review*, 55(3):231–244.

Hájek, J. (1970). A characterization of limiting distributions of regular estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 14(4):323–330.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic Press.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application*. Springer-Verlag New York.

Iacus, S. M. (2014). *sde: Simulation and Inference for Stochastic Differential Equations*. R package version 2.0.13. http://CRAN.R-project.org/package=sde.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation - Asymptotic Theory*. Springer-Verlag New York.

Jacobsen, M. (2001). Discretely observed diffusions: Classes of estimating functions and small Delta-optimality. *Scandinavian Journal of Statistics*, 28(1):123–149.

Jacobsen, M. (2002). Optimality and small Delta-optimality of martingale estimating functions. *Bernoulli*, 8(5):643–668.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Seminaire de Probabilites XXXI*, pages 232–246. Springer.

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics*, 27(1):83–96.

Jacod, J. (2006). Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli*, 12(3):383–401.

Jacod, J. (2010). Inference for stochastic processes. In Aït-Sahalia, Y. and Hansen, L. P., editors, *Handbook of Financial Econometrics: Applications*, volume 2 of *Handbooks in Finance*, chapter 16, pages 197–239. Elsevier.

Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *The Annals of Probability*, 26(1):267–307.

Jacod, J. and Protter, P. (2012). *Discretization of Processes*. Stochastic Modelling and Applied Probability. Springer-Verlag Berlin Heidelberg.

Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes*. Number 288 in Grundlehren der mathematischen Wissenschaften. Springer-Verlag, Berlin Heidelberg New York, 2nd edition.

Jacod, J. and Sørensen, M. (2012). Aspects of asymptotic statistical theory for stochastic processes. Preprint.

Jahn, P., Berg, R. W., Hounsgaard, J., and Ditlevsen, S. (2011). Motoneuron membrane potentials follow a time inhomogeneous jump diffusion process. *Journal of Computational Neuroscience*, 31(3):563–579.

Jakobsen, N. M. and Sørensen, M. (2015a). Efficient estimation for diffusions sampled at high frequency over a fixed time interval. Preprint.

Jakobsen, N. M. and Sørensen, M. (2015b). Efficient estimation for diffusions with jumps sampled at high frequency over an increasing time interval. Preprint.

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 173–212.

Kallenberg, O. (1997). *Foundations of Modern Probability*. Springer series in statistics: Probability and its applications. Springer.

Kawai, R. and Masuda, H. (2013). Local asymptotic normality for normal inverse Gaussian Lévy processes with high-frequency sampling. *ESAIM: Probability and Statistics*, 17:13–32.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, 24:211–229.

Kohatsu-Higa, A., Nualart, E., and Tran, N. K. (2014). LAN property for a simple Lévy process. *Comptes Rendus Mathematique*, 352(10):859–864.

Kohatsu-Higa, A., Nualart, E., and Tran, N. K. (2015). LAN property for a jump-diffusion process. Preprint.

Kou, S. G. (2002). A jump-diffusion model for option pricing. *Management Science*, 48(8):1086–1101.

Le Cam, L. (2002). The publications and writings of Lucien Le Cam. *The Annals of Statistics*, 30(3):683–687.

Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts*. Springer Series in Statistics. Springer New York, 2nd edition.

Li, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *The Annals of Statistics*, 41(3):1350–1380.

Mai, H. (2014). Efficient maximum likelihood estimation for Lévy-driven Ornstein-Uhlenbeck processes. *Bernoulli*, 20(2):919–957.

Mancini, C. (2004). Estimation of the characteristics of the jumps of a general Poisson-diffusion model. *Scandinavian Actuarial Journal*, 2004(1):42–52.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, 36(2):270–296.

Mancini, C. and Renò, R. (2011). Threshold estimation of Markov models with jumps and interest rate modeling. *Journal of Econometrics*, 160(1):77–92.

Masuda, H. (2007). Ergodicity and exponential beta-mixing bounds for multidimensional diffusions with jumps. *Stochastic Processes and their Applications*, 117(1):35–56.

Masuda, H. (2008). On stability of diffusions with compound-Poisson jumps. *Bulletin of Informatics and Cybernetics*, 40:61–74.

Masuda, H. (2011). Approximate quadratic estimating function for discretely observed Lévy driven SDEs with application to a noise normality test. *RIMS Kôkyûroku 1752*, pages 113–131.

Masuda, H. (2013). Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *The Annals of Statistics*, 41(3):1593–1641.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics*, 3(1-2):125–144.

Musila, M. and Lánský, P. (1991). Generalized Stein's model for anatomically complex neurons. *Biosystems*, 25(3):179–191.

Mykland, P. A. and Zhang, L. (2006). ANOVA for diffusions and Ito processes. *The Annals of Statistics*, 34(4):1931–1963.

Ogihara, T. and Yoshida, N. (2011). Quasi-likelihood analysis for the stochastic differential equation with jumps. *Statistical Inference for Stochastic Processes*, 14(3):189–229.

Papaspiliopoulos, O., Pokern, Y., Roberts, G. O., and Stuart, A. M. (2012). Nonparametric estimation of diffusions: a differential equations approach. *Biometrika*, 99(3):511–531.

Patel, A. and Kosko, B. (2008). Stochastic resonance in continuous and spiking neuron models with Lévy noise. *Neural Networks, IEEE Transactions on*, 19(12):1993–2008.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71.

Pedersen, A. R. (2000). Estimating the nitrous oxide emission rate from the soil surface by means of a diffusion model. *Scandinavian Journal of Statistics*, 27(3):385–403.

Picchini, U., Ditlevsen, S., and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology*, 25(2):141–155.

Podolskij, M. and Vetter, M. (2010). Understanding limit theorems for semimartingales: a short survey. *Statistica Neerlandica*, 64(3):329–351.

Prakasa Rao, B. L. S. (1983). Asymptotic theory for non-linear least squares estimator for diffusion processes. *Statistics: A Journal of Theoretical and Applied Statistics*, 14(2):195–209.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Rényi, A. (1963). On stable sequences of events. *Sankhyā: The Indian Journal of Statistics, Series A*, 25(3):293–302.

Roberts, G. O. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm. *Biometrika*, 88(3):603–621.

Schmisser, É. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics*, 47(1):61–84.

Schmisser, É. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stochastic Processes and their Applications*, 124(1):883–914.

Shimizu, Y. (2006a). Density estimation of Lévy measures for discretely observed diffusion processes with jumps. *Journal of The Japan Statistical Society*, 36(1):37–62.

Shimizu, Y. (2006b). M-estimation for discretely observed ergodic diffusion processes with infinitely many jumps. *Statistical Inference for Stochastic Processes*, 9(2):179–225.

Shimizu, Y. (2008). Some remarks on estimation of diffusion coefficients for jump-diffusions from finite samples. *Bulletin of Informatics and Cybernetics*, 40:51–60.

Shimizu, Y. (2009). Functional estimation for Lévy measures of semimartingales with Poissonian jumps. *Journal of Multivariate Analysis*, 100(6):1073–1092.

Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statistical Inference for Stochastic Processes*, 9:227–277.

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: a survey. *International Statistical Review*, 72(3):337–354.

Sørensen, M. (1991). Likelihood methods for diffusions with jumps. In Prabhu, N. U. and Basawa, I. V., editors, *Statistical Inference in Stochastic Processes*, pages 67–105. Marcel Dekker, New York.

Sørensen, M. (2010). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint.

Sørensen, M. (2012). Estimating functions for diffusion-type processes. In Kessler, M., Lindner, A., and Sørensen, M., editors, *Statistical Methods for Stochastic Differential Equations*, pages 1–107. Chapman & Hall.

Sørensen, M. and Uchida, M. (2003). Small-diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli*, 9(6):1051–1069.

Stramer, O., Bognar, M., and Schneider, P. (2010). Bayesian inference for discretely sampled Markov processes with closed-form likelihood expansions. *Journal of Financial Econometrics*, 8(4):450–480.

Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scandinavian Journal of Statistics*, 31(4):553–566.

Uchida, M. (2008). Approximate martingale estimating functions for stochastic differential equations with small noises. *Stochastic Processes and their Applications*, 118(9):1706–1721.

van der Meulen, F., Schauer, M., and van Zanten, H. (2014). Reversible jump MCMC for nonparametric drift estimation for diffusion processes. *Computational Statistics & Data Analysis*, 71:615–632.

van der Meulen, F. and van Zanten, H. (2013). Consistent nonparametric Bayesian inference for discretely observed scalar diffusions. *Bernoulli*, 19(1):44–63.

van der Vaart, A. (2002). The statistical work of Lucien Le Cam. *The Annals of Statistics*, 30(3):631–682.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, 41(2):220–242.

Yu, J. (2007). Closed-form likelihood approximation and estimation of jump-diffusions with an application to the realignment risk of the Chinese Yuan. *Journal of Econometrics*, 141(2):1245–1280.