Diffusion Models Observed at High Frequency and Applications in Finance

PhD thesis

Emil S. Jørgensen

Department of Mathematical Sciences
University of Copenhagen

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

Emil S. Jørgensen
Email: esj@math.ku.dk

Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
2100 Copenhagen
Denmark

Academic advisor:	Prof. Michael Sørensen, University of Copenhagen.
Assessment committee:	Prof. Arnaud Gloter, Université d'Évry Val d'Essonne; Prof. Helle Sørensen, University of Copenhagen; Dr. Almut Veraart, Imperial College London.
Submission date:	August 31, 2017.
ISBN:	978-87-7078-933-2.

Preface

The following thesis was submitted as part of the formal requirements to obtain a PhD degree from the University of Copenhagen. Most of my research was carried out at the Department of Mathematical Sciences, University of Copenhagen. Moreover, inspiring research stays at University of Technology Sydney and University of Chicago have had a significant impact on the final content.

First and foremost, I wish to express my sincere gratitude to my PhD advisor, Prof. Michael Sørensen. Your continuous support throughout the entire PhD program and, in particular, your mentoring role during my transition from applied mathematics to the frontier of asymptotic statistics for stochastic processes has been invaluable. Furthermore, I am grateful to Prof. Eckhard Platen for hosting me during my research stay at UTS Quantitative Finance Research Centre in Sydney and Prof. Per Mykland for inviting me to stay at the Stevanovich Center for Financial Mathematics, University of Chicago. The opportunity to discuss my research with each of you has been truly exciting.

A warm thank you also goes to my fellow PhD candidates and postdocs in Copenhagen. In particular, to Adam Lund and Frederik Mikkelsen for great times at the office; to Henrik Dam, Jannick Schreiner and Martin Jönsson for answering countless questions related to mathematical finance and great friendships on and off campus; and, finally, to Nina Jakobsen for stimulating conversations regarding our mutual research interests and excellent company at workshops around the world. A special thanks also to Bo Zhou for creating an enjoyable atmosphere in our shared office in Chicago.

Last but not least, my heartfelt thanks go to my closest family for believing in me and to Victoria for her endless love and support. Without your encouragement during the frustrating periods with high emotional volatility spikes, I wouldn't have been able to power through to the very end.

Emil S. Jørgensen

Resumé

Denne afhandling omhandler statistiske metoder til brug ved diffusionsmodeller inden for finansmatematik med diskrete observationer. Det primære fokus er på asymptotisk teori i forbindelse med højfrekvent data, navnligt inden for rammen af estimationsfunktioner. Ph.d.-afhandlingen består af 3 artikler, alle udarbejdet med henblik på udgivelse i et matematisk tidsskrift. Den første artikel er et studie af højfrekvent asymptotik for prædiktionsbaserede estimationsfunktioner med diskret-observerede diffusionsprocesser og er et samarbejde med min Ph.d.-vejleder, Michael Sørensen. Som hovedbidrag etablerer vi grænsesætninger for funktionaler af diffusionsprocesser der opfylder ρ-mixing-betingelsen og anvender resultaterne til at udlede eksistens af en konsistent, asymptotisk normal estimator for en bred klasse af prædiktionsbaserede estimationsfunktioner. Den anden artikel indeholder en udvidelse af den asymptotiske teori fra vores første artikel til det beslægtede tilfælde hvor vi observerer integralet af diffusionsprocessen. Udvidelsen baserer sig på udviklinger af funktionaler af diffusions- og integrerede diffusionsprocesser. Integrerede diffusioner er af naturlig interesse inden for finansiel matematik, hvor den observerbare volatilitet ofte anvendes til at filtrere en sti af den skjulte integrerede volatilitet. Den tredje og sidste artikel beskriver en parametrisk klasse af tidstransformerede diffusionsprocesser med henblik på at modellere risiko-spredte aktieindicer. Modellen er drevet af en enkelt Brownsk bevægelse, der repræsenterer den underliggende usikkerhed på markedet. Fokus er på relevante statistiske problemstillinger relateret til modelkonstruktionen og især behandler vi estimation af de ukendte modelparametre, samt konstruerer en simulations-baseret ikke-parametrisk test for den implicitte én-faktor hypotese for en bred klasse af kontinuerte Itô semimartingaler med stokastisk volatilitet.

Abstract

Broadly speaking, this thesis is devoted to statistical methods for discretely observed diffusion processes in finance. The main emphasis is on developing asymptotic theory for diffusions observed at high frequency, especially within the framework of estimating functions. The thesis consists of three papers, all intended for journal publication. The first paper is a study of high-frequency asymptotics for prediction-based estimating functions with discretely observed diffusion models and is joint work with my PhD advisor, Michael Sørensen. As our main contribution, we establish limit theorems for functionals of ρ-mixing diffusion processes and apply the results to derive existence of a consistent and asymptotically normal estimator for a tractable class of prediction-based estimating functions. The second paper contains an extension of our asymptotic results of the first paper to the case of discretely observed integrated diffusion processes. The extension relies on expansion results for functionals of diffusion and integrated diffusion processes. Integrated diffusions are of apparent interest in finance, where realized volatility or variations thereof are often used to construct a trajectory of the latent integrated volatility. The third and final paper deals with the construction of a parametric class of time-changed diffusion models aimed at modeling of diversified stock indices. The models are driven by a single Brownian motion that models the non-diversifiable risk of the underlying market. Emphasis is on relevant statistical problems related to the model construction and, in particular, we consider estimation of the parameters and construct a simulation-based nonparametric test for the implicit one-factor hypothesis for a large class of continuous Itô semimartingales with stochastic volatility.

Summary

The thesis consists of three research papers written during my PhD studies at University of Copenhagen in the time period from May 2014 until August 2017. Each paper can be read independently. Their respective titles and abstracts are as follows:

1. **Prediction-Based Estimation for Diffusion Models with High-Frequency Data.** This paper deals with prediction-based estimation for general, parametric diffusion models (X_t) with an unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$. We suppose that (X_t) is observed at equidistant time points $t_i^n = i\Delta_n$ for some $\Delta_n > 0$, and consider the ergodic high-frequency asymptotic scenario where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Subject to weak regularity conditions on (X_t), we prove existence of a consistent and asymptotically normal estimator $\hat{\theta}_n$ for a large class of prediction-based estimating functions. The proof of asymptotic normality requires the additional rate assumption $n\Delta_n^3 \to 0$. To complement the asymptotic results, we construct an explicit estimating function for the square-root (CIR) model and a simulation-based extension for the volatility process of a 3/2 model from finance. The latter illustrates a simple way to encompass diffusion models for which analytic moment conditions are not readily available into the estimation framework. The respective finite-sample behaviour of both estimating functions is evaluated based on simulated paths.

2. **Inference for Integrated Diffusions Observed at High Frequency.** In this paper, we suppose we observe a discretization $\{I_{t_i^n}\}_{i=0}^n$ of an integral process $I_t = \int_0^t X_s \, \mathrm{d}s$, where (X_t) is a time-homogeneous diffusion with an unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$ that we wish to estimate. The observation times $\{t_i^n\}$ are assumed to be deterministic and equidistant, i.e. $t_i^n = i\Delta_n$ for some $\Delta_n > 0$, and we consider the high-frequency asymptotic scenario where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Subject to mild regularity conditions on (X_t), we prove existence of a consistent and asymptotically normal estimator $\hat{\theta}_n$ for a tractable class of prediction-based estimating functions. The proofs are based on power expansions for diffusion and integrated diffusion models and asymptotic normality is obtained under the additional rate assumption $n\Delta_n^2 \to 0$. Our results are of particular interest in finance, where realized volatility or variations thereof are often used to construct a trajectory of the latent integrated volatility process.

3. **One-Factor Models for Diversified Stock Indices with High-Frequency Observations.** In this paper we construct a class of continuous-time stochastic volatility models aimed at modeling the dynamics of diversified stock indices. The models are of parametric diffusion-type and are driven by a single Brownian motion that models the non-diversifiable risk of the underlying market. For the construction we utilize the concept of stochastic market time and, in particular, the base process and the random time change are dependent processes in our setup. Our emphasis is on high-frequency econometric issues related to the model class. We propose a two-step method for estimating the finite-dimensional parameter and construct a simulation-based test for the implicit one-factor hypothesis for a large class of continuous Itô semimartingales with stochastic volatility. The one-factor test is based on a nonparametric measure of instantaneous leverage effect, where the one-factor model corresponds to perfect negative correlation. We illustrate the methodology using simulated data, as well as high-frequency observations of the S&P 500.

Contents

1 Introduction **1**

1.1 Discussion 1

1.2 Main contributions 4

2 Prediction-Based Estimation for Diffusion Models with High-Frequency Data **7**

2.1 Introduction 8

2.2 Preliminaries 9

2.2.1 Notation 10

2.2.2 Prediction-based estimating functions 11

2.2.3 Probabilistic notions 12

2.2.4 Assumptions 13

2.3 Limit theory for discretized diffusions 14

2.3.1 Law of large numbers 15

2.3.2 Potential operators for diffusion models 15

2.3.3 Central limit theory 16

2.4 Asymptotic theory 17

2.4.1 Simple predictor spaces 17

2.4.2 1-lag predictor spaces 18

2.4.3 Multiple predictor functions and optimal estimation 20

2.5 Estimating the asymptotic variance 22

2.5.1 Simple predictor spaces 22

2.5.2 1-lag predictor spaces 25

2.6 Applications in finance 26

2.7 Numerical results 29

2.8 Extensions and concluding remarks 32

2.9 Appendix A: Proofs 32
2.10 Appendix B: Moment expansions 44

3 Inference for Integrated Diffusions Observed at High Frequency 47
3.1 Introduction 48
3.2 Preliminaries 50
3.2.1 Notation 50
3.2.2 Model assumption 51
3.2.3 Prediction-based estimating functions 52
3.3 Euler-Itô expansions 54
3.3.1 Diffusion processes 54
3.3.2 Integrated diffusions 54
3.4 Limit theory for integrated diffusions 55
3.5 Asymptotic theory 56
3.5.1 Simple predictor spaces 57
3.5.2 1-lag predictor spaces 58
3.6 Extensions and concluding remarks 61
3.7 Appendix A: Proofs 61
3.8 Appendix B: Auxiliary results 83

4 One-Factor Models for Diversified Stock Indices with High-Frequency Observations 85
4.1 Introduction 86
4.2 One-factor index models 87
4.2.1 Time-change construction 87
4.2.2 Stochastic differential form 90
4.3 Model implications 91
4.3.1 Volatility 91
4.3.2 Continuous leverage effect 92
4.4 Parameter estimation 94
4.4.1 Market activity 94
4.4.2 Mean-reversion 95
4.5 Testing the one-factor hypothesis 97
4.6 Monte Carlo results 99

4.6.1 Path visualization . . . 99
4.6.2 Estimation . . . 100
4.6.3 Testing the one-factor hypothesis . . . 103
4.7 Empirical study: S&P 500 . . . 103
4.7.1 A closer look at the facts . . . 105
4.7.2 Index normalization . . . 107
4.8 Extensions and concluding remarks . . . 108

Bibliography **110**

1 Introduction

The amount of research literature devoted to statistical analysis of high-frequency data has exploded within the last few decades, driven by the availability of detailed transaction data from algorithmic trading in finance. The three papers that constitute this thesis fall within the interdisciplinary field of finance, probability theory and statistics commonly referred to as *high-frequency statistics* or *high-frequency financial econometrics*, depending on the particular research emphasis. Whereas the emblematic problems of high-frequency statistics are parametric and nonparametric estimation for discretely observed (discretized) continuous-time stochastic processes, the econometrics literature covers a wide class of filtering and testing problems of special relevance for high-frequency data in finance. Our main focus is on parametric problems for different types of diffusion models and the contributions are, in this sense, classical by nature.

1.1 Discussion

Diffusion processes are widely used in many scientific areas, particularly in finance. While the processes are characterized in terms of continuous-time dynamics, available time series are always observed at discrete points in time. To bridge the gap between the general theory of continuous-time stochastic processes and applications with discrete observations, statistical methods for discretely observed processes is an active area of research, and the recent availability of high-frequency data has spiked considerable interest into the construction of estimators and test statistics with nice asymptotic properties as the time between consecutive observations Δ_n goes to zero.

In the first paper (Chapter 2), we study parametric inference for discretely observed diffusion models (X_t) that satisfy a stochastic differential equation of the form

$$dX_t = a(X_t;\theta)dt + b(X_t;\theta)dB_t, \tag{1.1}$$

where (B_t) is standard Brownian motion and the unknown parameter of interest θ belongs to a subset $\Theta \subset \mathbb{R}^d$. Since the preferred method of maximum likelihood estimation is untractable for most diffusion models (1.1) applied in practice, a wide range of alternative methods have been proposed and applied successfully. The Markov property of (X_t) enables numerous types of quasi-likelihood, including contrast functions (Florens-Zmirou (1989), Yoshida (1992), Genon-Catalot and Jacod (1993), Hansen and Scheinkman (1995), Kessler (1997)), estimating functions (Bibby and M. Sørensen (1995), Kessler and M. Sørensen (1999), Kessler (2000), Jakobsen and M. Sørensen (2017)), likelihood expansions (Dacunha-Castelle and Florens-Zmirou (1986), Aït-Sahalia (2002), C. Li (2013)), Markov-chain Monte Carlo (Elerian, Chib, and Shephard (2001), Eraker (2001), Roberts and Stramer (2001)) and simulated likelihood (Beskos, Papaspiliopoulos, Roberts, and Fearnhead (2006), Beskos, Papaspiliopoulos, and Roberts (2009), Bladt and M. Sørensen (2014)). Complementing the purely parametric contributions, nonparametric estimation of the drift $a(\,\cdot\,)$ and the diffusion coefficient $b^2(\,\cdot\,)$ from discrete observation of (X_t) has been studied by Aït-Sahalia (1996a), Hansen, Scheinkman, and Touzi (1998), Hoffmann (1999a), Gobet, Hoffmann, and Reiß (2004) and Comte, Genon-Catalot, and Rozenholc (2007) under the assumption of strict stationarity of (X_t). Estimation for nonstationary, recurrent diffusion processes is considered by Bandi and Phillips (2003). With high-frequency observation of (X_t) on a *finite* time horizon $[0, T]$, estimation of the diffusion coefficient has been considered by Genon-Catalot, Larédo, and Picard (1992), Florens-Zmirou (1993), Hoffmann (1999a,b), Jacod (2000) and Renò (2008). A survey of nonparametric estimation methods with an extensive list of references is Fan (2005).

In the second paper (Chapter 3), we consider integrated diffusion models $(I_t)_{t\geq 0}$ of the general form

$$\begin{aligned} dI_t &= X_t dt \\ dX_t &= a(X_t;\theta)dt + b(X_t;\theta)dB_t, \end{aligned}$$

where (X_t) takes values in an open interval $(l, r) \subset \mathbb{R}$ and once again the parameter of interest $\theta \in \Theta \subset \mathbb{R}^d$. Although to a lesser extent than for discretely observed diffusions of the form (1.1), parametric estimation for integrated diffusions has also been the topic of many papers in econometrics and statistics, the former in the guise of continuous-time stochastic volatility models. If we consider the simple stochastic volatility model

$$dS_t = \sqrt{v_t}dW_t, \tag{1.2}$$

where (W_t) denotes a standard Brownian motion, the availability of high-frequency data generated by (S_t) enables us to filter a trajectory of the latent integrated volatility

$$\int_0^t v_s \,\mathrm{d}s \tag{1.3}$$

and view it as an observable process, possibly with measurement error. This has lead to the construction of estimators for integrated processes in the case where $v_t = v_t(\theta)$ for a parameter $\theta \in \Theta \subset \mathbb{R}^d$, e.g. if the volatility dynamics itself is described by a time-homogeneous diffusion process

$$dv_t = \mu(v_t;\theta)dt + \sigma(v_t;\theta)dB_t.$$

Known diffusion models for the volatility dynamics include the GARCH(1,1) diffusion model in Nelson (1990), the square-root (CIR) process in Heston (1993) and the 3/2 diffusion in Drimus (2012). Parametric estimation based on realized power variations that approximate the integrated volatility (1.3) have been studied by e.g. Bollerslev and Zhou (2002), Barndorff-Nielsen and Shephard (2002a) and Todorov (2009), the latter in a general method-of-moment (GMM) framework for a large class of stochastic volatility models with jumps. On a related note, the recent paper by J. Li and Xiu (2016) appears to be the first to develop high-frequency (infill) asymptotics for GMM estimators of parameters in the diffusion coefficient of the volatility process by preliminary filtering of the spot volatility instead. On the statistical side, Ditlevsen and M. Sørensen (2004) illustrate a simple way to construct explicit Godambe optimal prediction-based estimating functions for (X_t) belonging to the tractable class of Pearson diffusions defined in Forman and M. Sørensen (2008), and Baltazar-Larios and M. Sørensen (2010) propose a simulated EM-algorithm to obtain maximum likelihood estimators for integrated diffusions contaminated by noise. A third, and for our paper highly influential, approach based on expansion results for small values of Δ_n was proposed by Gloter (2000, 2006). His construction of contrast estimators utilizes the basic observation that, as $\Delta_n \to 0$,

$$\Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} X_s \, \mathrm{d}s \approx X_{(i-1)\Delta_n}$$

which allows for the derivation of high-frequency limit theorems for integrated diffusions. Non-parametric estimation of the drift $a(\,\cdot\,)$ and diffusion coefficient $b^2(\,\cdot\,)$ from high-frequency observations of (I_t) was studied by Comte, Genon-Catalot, and Rozenholc (2009).

Lastly, building on recent work by Ignatieva and Platen (2012) and Platen and Rendek (2012b), the third paper (Chapter 4) considers a class of time-changed square-root diffusion processes $S_t = X_{\tau_t}$, where the base process (X_t) is defined as a strong solution of the stochastic differential equation

$$dX_t = \beta(1 - X_t)dt + \sqrt{X_t}dW_t$$

for a carefully constructed Brownian motion (W_t). We model the time evolution $\tau = (\tau_t)$ using the mathematical concept of time change, specifically by letting τ be an integral w.r.t. a latent market activity process (M_t) that takes values in $(0, \infty)$, i.e. we define

$$\tau_t(\omega) = \int_0^t M_s(\omega) \, \mathrm{d}s. \tag{1.4}$$

The use of time change as a means to construct stochastic volatility models in finance is widely accepted; see e.g. A. Veraart and Winkel (2010) for a concise review. Notably, the idea of replacing t-time by a stochastic market time goes back to Clark (1973), who was the first to model asset price dynamics using time-changed Brownian motion. His approach was extended by Ané and Geman (2000) to include a more general definition of market time. The use of time-changes obtained by integrating over a positive stochastic process as in (1.4) has been studied by e.g. Carr et al. (2003) for the construction of Lévy processes with stochastic volatility. Whereas many continuous-time processes with a natural time-change representation are defined in such a way that the base process (X_t) and the time change τ are independent, our construction of (W_t) by means of the Dubins-Schwarz representation for continuous local martingales implies that they are dependent processes in our setup. Examples where the time change is independent of the base process includes Clark (1973), Madan and Seneta (1990) and Barndorff-Nielsen (1997) for Lévy processes, or Ané and Geman (2000) and Carr et al. (2003) for models with stochastic volatility.

1.2 Main contributions

In the following we briefly describe the individual contribution of each paper in more mathematical detail and emphasize how the results relate to the general discussion above.

In the first paper, we let (X_t) take values in an open interval $(l, r) \subset \mathbb{R}$ and suppose that the process is stationary under the probability measure $\mathbb{P}_\theta$ for an invariant initial distribution $X_0 \sim \mu_\theta$. For estimation purposes we observe a single discretization

$$X_0, X_{t_1^n}, \ldots, X_{t_n^n},$$

and we assume that the observation times are deterministic and equidistant and write $t_i^n = i\Delta_n$ for the appropriate $\Delta_n > 0$. To encompass consistent estimation of both drift and diffusion parameters into the asymptotic theory, we consider the ergodic high-frequency sampling scenario

$$n \to \infty, \quad \Delta_n \to 0, \quad n \cdot \Delta_n \to \infty, \tag{1.5}$$

where the time horizon $T_n = n\Delta_n$ increases with the number of observations. The construction of estimators $\hat{\theta}_n$ is carried out in the framework of prediction-based estimating functions. These estimating functions were proposed by M. Sørensen (2000, 2011) as a versatile estimation framework for non-Markovian diffusion-type models and generalize the martingale estimating functions pioneered by Bibby and M. Sørensen (1995). We show that, with a restriction to finite-dimensional predictor spaces, the estimating functions of this paper lie outside the class of approximate martingale estimating functions defined in M. Sørensen (2017), but still yield consistent and asymptotically normal estimators under mild regularity conditions on (X_t). Asymptotic normality requires the additional rate assumption $n\Delta_n^3 \to 0$. Whereas the use of prediction-based estimating functions for inference in diffusion models (1.1) comes at a loss of efficiency compared to other estimation methods, the asymptotic results of this paper provide the foundation for our study of (non-Markovian) integrated diffusions observed at high frequency in Chapter 3.

The second paper extends the asymptotic results for prediction-based estimating functions of the first paper to the related observational scheme of discretely observed integrated diffusions $(I_t)_{t\geq 0}$ of the form

$$\begin{aligned} dI_t &= X_t dt \\ dX_t &= a(X_t; \theta)dt + b(X_t; \theta)dB_t. \end{aligned}$$

We observe a single discretization $\{I_{t_i^n}\}_{i=0}^n$ of the integrated process at deterministic, equidistant points in time and once again consider the ergodic high-frequency sampling scenario

(1.5). The latent diffusion process (X_t) is strictly stationary under the probability measure $\mathbb{P}_\theta$ for an invariant distribution $X_0 \sim \mu_\theta$. A more appropriate, *equivalent* observation scheme is obtained for the transformed variables

$$Y_i = \Delta_n^{-1}\left(I_{t_i^n} - I_{t_{i-1}^n}\right) = \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} X_s\,\mathrm{d}s \tag{1.6}$$

where $i = 1, \ldots, n$. Whereas low-frequency asymptotic results for prediction-based estimating functions with integrated observations follow from general results in M. Sørensen (2000), high-frequency asymptotics require quite a bit of effort. Our proofs rely on preliminary functional versions of the classic Euler approximation

$$X_{t_i^n}\,|\,\mathcal{F}_{t_{i-1}^n} \approx \mathcal{N}\left(X_{t_{i-1}^n} + \Delta_n a(X_{t_{i-1}^n};\theta), \Delta_n b^2(X_{t_{i-1}^n};\theta)\right)$$

and the similar result that

$$Y_i\,|\,\mathcal{F}_{t_{i-1}^n} \approx \mathcal{N}\left(X_{t_{i-1}^n} + \Delta_n \frac{1}{2} a(X_{t_{i-1}^n};\theta), \Delta_n \frac{1}{3} b^2(X_{t_{i-1}^n};\theta)\right)$$

for small values of Δ_n, combined with the asymptotic theory for diffusion models developed in Chapter 2. We show that, under suitable regularity conditions, consistency and asymptotic normality of prediction-based estimators $\hat{\theta}_n$ is once again attained within the ergodic scenario (1.5). Asymptotic normality requires the strong additional rate assumption $n\Delta_n^2 \to 0$.

Our emphasis in the final paper is on high-frequency econometric issues related to the index model (S_t). For estimation purposes, we show that we can view (S_t) as the observable marginal of a bivariate diffusion model

$$\begin{aligned} dS_t &= \beta(1-S_t)M_t dt + \sqrt{S_t M_t} dB_t \\ dM_t &= a(M_t)dt + b(M_t)dB_t, \end{aligned}$$

where (M_t) denotes the market activity process in (1.4). We propose a two-step method for estimating the mean-reversion parameter $\beta > 0$ in the dynamics of (S_t), as well as any parameter $\gamma \in \Gamma \subset \mathbb{R}^d$ that appears in (M_t). Initially, we demonstrate how to filter a trajectory of (M_t) or τ using a Lamperti transformation which, in turn, enables us to estimate γ using any of standard methods for time-homogeneous diffusion processes described in Section 1.1. Secondly, we deal with the estimation of β which, despite our ability to filter (M_t) or, equivalently, τ from the discretization of (S_t), remains complicated due to the dependence between the base process (X_t) and τ. Our proposed estimator exploits the common driving Brownian motion (B_t) to construct an explicit estimating equation which can be solved for β. Finally, we construct a simulation-based test for the implicit one-factor hypothesis for a large class of continuous Itô semimartingales with stochastic volatility. The test is fully nonparametric and based on a normalized measure of instantaneous leverage effect proposed by Kalnina and Xiu (2017), where the one-factor model corresponds to perfect negative correlation.

2

Prediction-Based Estimation for Diffusion Models with High-Frequency Data

Emil S. Jørgensen and Michael Sørensen
University of Copenhagen

Abstract. Prediction-based estimating functions provide a versatile framework for parameter estimation in discretized diffusion-type models. This paper deals with prediction-based estimation for general, parametric diffusion models (X_t) with an unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$. We suppose that (X_t) is observed at equidistant time points $t_i^n = i\Delta_n$ for some $\Delta_n > 0$, and consider the ergodic high-frequency asymptotic scenario where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Subject to weak regularity conditions on (X_t), we prove existence of a consistent and asymptotically normal estimator $\hat{\theta}_n$ for a large class of prediction-based estimating functions. The proof of asymptotic normality requires the additional rate assumption $n\Delta_n^3 \to 0$. To complement the asymptotic results, we construct an explicit estimating function for the square-root (CIR) model and a simulation-based extension for the volatility process of a 3/2 model from finance. The latter illustrates a simple way to encompass diffusion models for which analytic moment conditions are not readily available into the estimation framework. The respective finite-sample behaviour of both estimating functions is evaluated based on simulated paths.

Keywords: *Diffusion process, high-frequency data, infinitesimal generator, potential operator, prediction-based estimating functions, ρ-mixing.*

2.1 Introduction

Diffusion processes are often used to model the behaviour of stochastic dynamical systems, especially in finance. These processes are characterized in terms of probabilistic behaviour in continuous time, but for most applications we only observe the system at discrete points in time. The ability to fit a particular diffusion model to a discrete set of observations is crucial if we are to, e.g., predict future values of a given time series, and statistical methods for discretely observed (discretized) stochastic processes is an active area of research. In particular, the recent availability of high-frequency data has generated considerable interest in the asymptotic behaviour of estimators and test statistics as the time between consecutive observations goes to zero.

In this paper, we study parametric inference for diffusion models that satisfy a stochastic differential equation of the form

$$dX_t = a(X_t;\theta)dt + b(X_t;\theta)dB_t, \tag{2.1}$$

where (B_t) is standard Brownian motion and the parameter of interest θ belongs to a subset $\Theta \subset \mathbb{R}^d$. We suppose that (X_t) takes values in an open interval $(l, r) \subset \mathbb{R}$ and is stationary under the probability measure $\mathbb{P}_\theta$ for an invariant initial distribution $X_0 \sim \mu_\theta$. For estimation purposes we observe a single discretization

$$X_0, X_{t_1^n}, \dots, X_{t_n^n},$$

and we assume that the observation times are deterministic and equidistant, i.e. $t_i^n = i\Delta_n$ for some $\Delta_n > 0$. To encompass consistent estimation of both drift and diffusion parameters into the asymptotic theory, we consider the ergodic high-frequency sampling scenario

$$n \to \infty, \quad \Delta_n \to 0, \quad n \cdot \Delta_n \to \infty, \tag{2.2}$$

where the time horizon $T_n = n\Delta_n$ increases with the number of observations.

The construction of estimators $\hat{\theta}_n$ is carried out within the framework of prediction-based estimating functions. These estimating functions were proposed by M. Sørensen (2000, 2011) as a versatile estimation framework for non-Markovian diffusion-type models and generalize the martingale estimating functions pioneered by Bibby and M. Sørensen (1995). We show that, with a restriction to finite-dimensional predictor spaces, the estimating functions of this paper lie outside the class of approximate martingale estimating functions defined in M. Sørensen (2017), but still lead to both consistent and asymptotically normal estimators under mild regularity conditions on (X_t). The latter requires the additional rate assumption $n\Delta_n^3 \to 0$. Whereas the use of prediction-based estimating functions for diffusion models (2.1) comes at a loss of efficiency compared to other estimation methods, the asymptotic results of this paper provide the foundation for our study of (non-Markovian) integrated diffusions observed at high frequency in Chapter 3.

Parametric estimation for discretely observed diffusion processes has been the topic of many papers in econometrics and statistics. Since the preferred method of exact maximum likelihood is untractable for most diffusion models applied in practice, a wide range of alternative methods have been proposed and applied successfully. The Markov property of (X_t) enables many types of quasi-likelihood, including contrast functions (Florens-Zmirou (1989), Yoshida (1992), Genon-Catalot and Jacod (1993), Hansen and Scheinkman (1995), Kessler (1997)), estimating functions (Bibby and M. Sørensen (1995), Kessler and M. Sørensen (1999), Kessler (2000), Jakobsen and M. Sørensen (2017)), likelihood expansions (Dacunha-Castelle and Florens-Zmirou (1986), Aït-Sahalia (2002), C. Li (2013)), Markov-chain Monte Carlo (Elerian, Chib, and Shephard (2001), Eraker (2001), Roberts and Stramer (2001)) and simulated likelihood (Beskos, Papaspiliopoulos, Roberts, and Fearnhead (2006), Beskos, Papaspiliopoulos, and Roberts (2009), Bladt and M. Sørensen (2014)).

Complementing the parametric literature, nonparametric estimation of the drift $a(\,\cdot\,)$ and the diffusion (volatility) coefficient $b^2(\,\cdot\,)$ from discrete observation of (X_t) has been studied by Aït-Sahalia (1996a), Hansen, Scheinkman, and Touzi (1998), Hoffmann (1999a), Gobet, Hoffmann, and Reiß (2004) and Comte, Genon-Catalot, and Rozenholc (2007) under the assumption of strict stationarity of (X_t). Estimation for nonstationary, recurrent diffusion processes is considered by Bandi and Phillips (2003). With high-frequency observation of (X_t) on a *finite* time horizon $[0, T]$, estimation of the diffusion coefficient has been considered by Genon-Catalot, Larédo, and Picard (1992), Florens-Zmirou (1993), Hoffmann (1999a,b), Jacod (2000) and Renò (2008). An excellent survey of nonparametric estimation methods with an extensive list of references is Fan (2005).

The structure of the paper is as follows. In Section 2.2 we present the general notation used throughout the paper, define a tractable class of prediction-based estimating functions, and formulate our general assumption on (X_t) for the asymptotic theory. Section 2.3 is devoted to limit theorems for functionals

$$V_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_{t_{i-1}^n})$$

and, in particular, we establish a Gaussian central limit theorem (CLT) for f belonging to a large class of functions. The variance of the Gaussian limit involves the potential of f which is characterized as the solution of a probabilistic Poisson-type differential equation. Asymptotic results are provided in Section 2.4. In Section 2.5 we construct Monte Carlo-based estimators for the asymptotic variances obtained in Section 2.4. To complement the asymptotic results, we construct an explicit estimating function for the square-root process of Cox, Ingersoll, and Ross (1985) and a simulation-based extension for the volatility process of a 3/2 model from finance. Numerical results are provided in Section 2.7. Section 2.8 concludes. All proofs are deferred to Section 2.9 and Section 2.10 contains some auxiliary results applied in the proofs.

2.2 Preliminaries

In this section we introduce the notation used throughout the paper, define a tractable class of prediction-based estimating functions, recall some core notions from probability theory, and formulate our main assumptions on the diffusion model (X_t) and the parameter space Θ for the asymptotic theory.

2.2.1 Notation

Aligned with the econometric and statistical literature on parametric inference for stochastic processes, our general notation is as follows:

1. The parameter of interest $\theta \in \Theta \subset \mathbb{R}^d$ for $d \geq 1$. We denote the true parameter by θ_0.
2. We denote the state space of X by $(S, \mathscr{B}(S))$ and assume throughout that S is an open interval in $\mathbb{R}$, i.e. $S = (l, r)$ for $-\infty \leq l < r \leq \infty$, endowed with its Borel σ-algebra $\mathscr{B}(S)$.
3. For short, we write $\mu_\theta(f) = \int_S f(x)\mu_\theta(dx)$ for functions $f : S \to \mathbb{R}$ and denote by $\mathscr{L}^p(\mu_\theta)$ the space of equivalence classes of p-integrable functions w.r.t. the invariant measure μ_θ. In particular, $\|f\|_2$ denotes the canonical norm on $\mathscr{L}^2(\mu_\theta)$ defined by

$$\|f\|_2 = \left(\int_S f^2(x)\mu_\theta(dx)\right)^{1/2}.$$

4. For random variables Y and Z defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we write $Y \leq_C Z$ if there exists a constant $C > 0$ such that $Y \leq C \cdot Z$, $\mathbb{P}$-almost surely.

To define some function spaces of interest, we say that $f : S \times \Theta \to \mathbb{R}$ is of *polynomial growth in x* if for every $\theta \in \Theta$, $|f(x;\theta)| \leq_C 1 + |x|^C$ for $x \in S$.

5. We denote by $\mathcal{C}_p^{j,k}(S \times \Theta)$, $j, k \geq 0$, the class of real-valued functions $f(x;\theta)$ satisfying that

   - f is j times continuously differentiable w.r.t. x;
   - f is k times continuously differentiable w.r.t. $\theta_1, \ldots, \theta_d$;
   - f and all partial derivatives $\partial_x^{j_1}\partial_{\theta_1}^{k_1}\cdots\partial_{\theta_d}^{k_d} f$, $j_1 \leq j$, $k_1 + \cdots + k_d \leq k$, are of polynomial growth in x.

   Similarly, we define $\mathcal{C}_p^j(S)$.

6. For use in the appendices, $R(\Delta, x;\theta)$ denotes a generic function such that

$$|R(\Delta, x;\theta)| \leq_C F(x;\theta), \tag{2.3}$$

   where F is of polynomial growth in x. We sometimes write $R_0(\Delta, x;\theta)$ to emphasize that the remainder term $R(\Delta, x;\theta)$ also depends on the true parameter θ_0.

2.2.2 Prediction-based estimating functions

The general theory of prediction-based estimating functions was developed by M. Sørensen (2000) and later reviewed and extended in M. Sørensen (2011). In this paper we consider estimating functions of the general form

$$G_n(\theta) = \sum_{i=q}^{n} \sum_{j=1}^{N} \pi_{i-1,j} \left[f_j(X_{t_i^n}) - \breve{\pi}_{i-1,j}(\theta) \right], \tag{2.4}$$

where $\{f_j\}_{j=1}^N$ is a finite set of real-valued functions in $\mathscr{L}^2(\mu_\theta)$ and for every $j \in \{1,\ldots,N\}$, $\breve{\pi}_{i-1,j}(\theta)$ denotes the orthogonal $\mathscr{L}^2(\mu_\theta)$-projection of $f_j(X_{t_i^n})$ onto a finite-dimensional subspace

$$\mathcal{P}_{i-1,j} = \operatorname{span}\left\{1, f_j\left(X_{t_{i-1}^n}\right), \ldots, f_j\left(X_{t_{i-q_j}^n}\right)\right\} \subset \mathscr{L}^2(\mu_\theta) \tag{2.5}$$

for a fixed $q_j \geq 0$. The coefficients $\pi_{i-1,j}$ are d-dimensional column vectors with entries belonging to $\mathcal{P}_{i-1,j}$.

The collection of subspaces $\{\mathcal{P}_{i-1,j}\}_{ij}$ are referred to as *predictor spaces*. In this sense, what we predict are values of $f_j(X_{t_i^n})$ for each $i \geq q := \max_{1\leq j\leq N} q_j$. Most prediction-based estimating functions applied in practice are of this particular form; see e.g. M. Sørensen (2000) for applications to discretized stochastic volatility models, and Ditlevsen and M. Sørensen (2004) for the case of integrated diffusions.

Since the predictor space $\mathcal{P}_{i-1,j}$ is closed, the $\mathscr{L}^2(\mu_\theta)$-projection of $f_j(X_{t_i^n})$ onto $\mathcal{P}_{i-1,j}$ is well-defined and uniquely determined by the normal equations

$$\mathbb{E}_\theta\left(\pi\left[f_j(X_{t_i^n}) - \breve{\pi}_{i-1,j}(\theta)\right]\right) = 0, \tag{2.6}$$

for all $\pi \in \mathcal{P}_{i-1,j}$; see e.g. Rudin (1987). Here and in everything that follows, $\mathbb{E}_\theta(\,\cdot\,)$ denotes expectation w.r.t. the underlying probability measure $\mathbb{P}_\theta$. Moreover, by restricting ourselves to predictor spaces of the form (2.5), as well as only diffusion models (X_t) that are stationary under $\mathbb{P}_\theta$, the orthogonal projection $\breve{\pi}_{i-1,j}(\theta) = \breve{a}_n(\theta)_j^T Z_{i-1,j}$ where

$$Z_{i-1,j} = \left(1, f_j\left(X_{t_{i-1}^n}\right), \ldots, f_j\left(X_{t_{i-q_j}^n}\right)\right)^T \tag{2.7}$$

and $\breve{a}_n(\theta)_j^T$ is the unique (q_j+1)-dimensional coefficient vector

$$\breve{a}_n(\theta)_j^T = \left(\breve{a}_n(\theta)_{j0}, \breve{a}_n(\theta)_{j1} \ldots, \breve{a}_n(\theta)_{jq_j}\right)$$

determined by the moment conditions

$$\mathbb{E}_\theta\left[Z_{q_j-1,j} f_j\left(X_{t_{q_j}^n}\right)\right] - \mathbb{E}_\theta\left[Z_{q_j-1,j} Z_{q_j-1,j}^T\right] \breve{a}_n(\theta)_j = 0. \tag{2.8}$$

Note that in the simplest case where $q_j = 0$, $\mathcal{P}_{i-1,j} = \operatorname{span}\{1\}$ and it follows immediately from the normal equations (2.6) that $\breve{\pi}_{i-1,j}(\theta) = \mu_\theta(f_j)$.

We obtain an estimator $\hat{\theta}_n$ by solving the estimating equation

$$G_n(\theta) = 0$$

and often refer to $\hat{\theta}_n$ as a G_n*-estimator*.

Remark 2.2.1. If we define an equivalence relation $\sim$ on the set of estimating functions of the form (2.4) by $G_n \sim H_n$ *if and only if* $H_n = M_n G_n$ for an invertible $d \times d$-matrix M_n, equivalent estimating functions yield identical estimators $\hat{\theta}_n$. In particular, estimators obtained from equivalent estimating functions share the same asymptotic properties. We freely apply this property in the proofs of Section 2.4.

2.2.3 Probabilistic notions

Two notions from the theory of stochastic processes play a central role in this paper; the infinitesimal generator of a diffusion process (X_t), and the dependence property known as ρ-mixing. We briefly introduce both concepts below.

Infinitesimal generator With any weak solution of (2.1), we associate a family of operators $(P_t^\theta)_{t\geq 0}$ where for $f \in \mathscr{L}^1(\mu_\theta)$,

$$P_t^\theta f(x) = \mathbb{E}_\theta\left(f(X_t) \mid X_0 = x\right),$$

and the operator that maps a function f onto $P_t^\theta f$ is known to satisfy a number of useful properties; see e.g. Hansen and Scheinkman (1995). In particular, $P_t^\theta : \mathscr{L}^2(\mu_\theta) \to \mathscr{L}^2(\mu_\theta)$ and the semigroup property $P_t^\theta \circ P_s^\theta = P_{t+s}^\theta$ holds for all $t, s \geq 0$.

The *(infinitesimal) generator* $\mathcal{A}_\theta$ of a diffusion (X_t) is defined as the limiting operator

$$\mathcal{A}_\theta f = \lim_{t\to 0} \frac{P_t^\theta f - f}{t},$$

whenever the limit $\mathcal{A}_\theta f$ is well-defined in $\mathscr{L}^2(\mu_\theta)$. The *domain* of $\mathcal{A}_\theta$ is the collection $\mathcal{D}_{\mathcal{A}_\theta}$ of all elements $f \in \mathscr{L}^2(\mu_\theta)$ for which $\mathcal{A}_\theta f$ exists.

For diffusion processes there is a well-known connection between the generator $\mathcal{A}_\theta$ and the drift and diffusion coefficients of (X_t); using Itô's formula, it can be shown subject to regularity conditions on (X_t) that $\mathcal{A}_\theta f = \mathcal{L}_\theta f$, where

$$\mathcal{L}_\theta f(x) = a(x;\theta)\partial_x f(x) + \frac{1}{2} b^2(x;\theta)\partial_x^2 f(x) \tag{2.9}$$

for f belonging to a large subset of the domain $\mathcal{D}_{\mathcal{A}_\theta}$. For our purposes it suffices to note that whenever (X_t) satisfies Condition 2.2.2 below, $\mathcal{C}_p^2(S) \subset \mathcal{D}_{\mathcal{A}_\theta}$ and the explicit representation $\mathcal{A}_\theta f = \mathcal{L}_\theta f$ holds for all $f \in \mathcal{C}_p^2(S)$; see e.g. Kessler (2000).

Recall that $\lambda \in \mathbb{R}$ is an eigenvalue of $\mathcal{A}_\theta$ if

$$\mathcal{A}_\theta f = \lambda f$$

for some $f \in \mathcal{D}_{\mathcal{A}_\theta}$. The collection of all eigenvalues is known as the *spectrum* of $\mathcal{A}_\theta$ and will be denoted by $\mathscr{S}(\mathcal{A}_\theta)$. From spectral theory it is known that $\mathscr{S}(\mathcal{A}_\theta) \subset (-\infty, 0]$. If $\mathscr{S}(\mathcal{A}_\theta) \subset (-\infty, -\lambda^*] \cup \{0\}$ for some $\lambda^* > 0$, the generator $\mathcal{A}_\theta$ is said to have a *spectral gap*. In particular, whenever the diffusion process (X_t) is ergodic and reversible under $\mathbb{P}_\theta$, the existence of a spectral gap $\lambda^* > 0$ is equivalent to (X_t) satisfying the ρ-mixing property; see Genon-Catalot, Jeantheau, and Larédo (2000). In Section 2.3 we apply this equivalence to establish the existence of the potential $U_\theta(f)$ for a large class of functions f.

Mixing For general stochastic processes, mixing coefficients provide a way of measuring how dynamic dependence decays over time. Various notions, including α-, β- and ρ-mixing, appear in the literature and are often used to establish central limit theorems for processes that are not martingales; see e.g. Doukhan (1994).

For a general probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and σ-algebras $\mathcal{A}, \mathcal{B} \subset \mathcal{F}$, the ρ-mixing coefficient

$$\rho(\mathcal{A}, \mathcal{B}) := \sup \left\{ |\mathrm{Corr}(X, Y)| : X \in \mathscr{L}^2(\Omega, \mathcal{A}, \mathbb{P}), Y \in \mathscr{L}^2(\Omega, \mathcal{B}, \mathbb{P}) \right\},$$

where X and Y are random variables having values in $\mathbb{R}$. Hence, by considering the σ-algebras $\mathcal{F}_t := \sigma(X_s : s \leq t)$ and $\mathcal{F}^t := \sigma(X_s : s > t)$ generated by a stochastic process (X_t) defined on the same space, we construct a dynamic measure of dependence

$$\rho_X(t) = \sup_{s \geq 0} \rho\left(\mathcal{F}_s, \mathcal{F}^{t+s}\right),$$

which, for a stationary Markov process, takes the simple form

$$\rho_X(t) = \rho\left(\sigma(X_0), \sigma(X_t)\right). \tag{2.10}$$

A process (X_t) is said to be *ρ-mixing* if $\rho_X(t) \to 0$ as $t \to 0$. A review of mixing properties for stationary Markov processes can be found in Genon-Catalot, Jeantheau, and Larédo (2000).

2.2.4 Assumptions

To derive asymptotic results for diffusion models of the general form (2.1), we impose some mild dependence and regularity conditions on (X_t). In particular, the following condition is sufficient to establish limit theorems for functionals

$$\frac{1}{n} \sum_{i=1}^{n} f(X_{t_{i-1}^n}),$$

for $f : S \to \mathbb{R}$ belonging to a suitable class of functions. We return to this topic in Section 2.3.

Condition 2.2.2. *For any $\theta \in \Theta$, the stochastic differential equation*

$$dX_t = a(X_t;\theta)dt + b(X_t;\theta)dB_t, \quad X_0 \sim \mu_\theta$$

has a weak solution $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}_\theta, (B_t), (X_t))$ *with the property that*

- (X_t) *is stationary, ρ-mixing under $\mathbb{P}_\theta$.*

Moreover, the a priori given triplet (a, b, μ_θ) satisfies the regularity conditions

- $a, b \in \mathcal{C}_p^{2,0}(S \times \Theta)$,
- $|a(x;\theta)| + |b(x;\theta)| \leq_C 1 + |x|$,
- $b(x;\theta) > 0$ *for* $x \in S$,
- $\int_S |x|^k \mu_\theta(dx) < \infty$ *for all* $k \geq 1$.

The reader can consult e.g. Kallenberg (2002) for a concise introduction to weak solutions of stochastic differential equations. For the discretized filtration $\{\mathcal{F}_{t_i^n}\}$, we let $\mathcal{F}_i^n := \mathcal{F}_{t_i^n}$ and the notation $\mu_0 = \mu_{\theta_0}$, $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$, etc., is applied throughout the paper.

Of course, we want Condition 2.2.2 to hold for the true, but unknown, parameter θ_0. A conventional convexity restriction on Θ ensures that we obtain asymptotic normality of our prediction-based estimators $\hat{\theta}_n$ in Section 2.4 from standard arguments.

Condition 2.2.3. *The parameter $\theta \in \Theta \subset \mathbb{R}^d$ and it holds that*

- $\theta_0 \in \operatorname{int}(\Theta)$,
- Θ *is convex.*

Here $\operatorname{int}(\Theta)$ denotes the interior of Θ, i.e. the union of all open sets contained in Θ.

2.3 Limit theory for discretized diffusions

This section is devoted to limit theorems for functionals

$$V_n(f) = \frac{1}{n} \sum_{i=1}^{n} f(X_{t_{i-1}^n}), \tag{2.11}$$

where f takes values in $\mathbb{R}$ and $\{X_{t_i^n}\}_{i=0}^n$ is a discretization of a diffusion process (X_t) that satisfies Condition 2.2.2. In Section 2.3.1 we state a law of large numbers for a large class of functions f. The CLT requires a bit more caution and stronger regularity assumptions on f. We define a suitable class of functions $\mathscr{H}_\theta^2$ in Section 2.3.2 and formulate and prove a CLT in Section 2.3.3. The results are given w.r.t. the true probability measure $\mathbb{P}_0$.

2.3.1 Law of large numbers

For the asymptotic theory in Section 2.4, it suffices to establish pointwise convergence of $V_n(f)$. The following result follows from the continuous-time ergodic theorem; see Corollary 10.9, Kallenberg (2002).

Lemma 2.3.1. *Let $f \in \mathcal{C}_p^1(S)$. Then,*

$$V_n(f) \xrightarrow{\mathbb{P}_0} \mu_0(f).$$

Lemma 2.3.1 will be applied frequently in proofs later on. Conversely, the only application of our subsequent CLT (Proposition 2.3.4) in this paper is to establish asymptotic normality of G_n-estimators in Section 2.4. This enables us to restrict attention to a smaller, more appropriate class of functions f; indeed, the normal equations (2.6) imply that any prediction-based estimating function of the form (2.4) is unbiased, i.e.

$$\mathbb{E}_\theta\left(G_n(\theta)\right) = 0, \tag{2.12}$$

and, correspondingly, we only consider functions $f : S \to \mathbb{R}$ for which $\mu_\theta(f) = 0$ for the remainder of this section.

2.3.2 Potential operators for diffusion models

The variance of the Gaussian limit distribution in Proposition 2.3.4 involves the potential of the function f which we define as

$$U_\theta(f)(x) = \int_0^\infty P_t^\theta f(x)\,\mathrm{d}t \tag{2.13}$$

for an implicit choice of diffusion model (X_t).

By construction, the potential (2.13) has a natural operator interpretation. To identify a domain and codomain for the operator $f \mapsto U_\theta(f)$, we apply that the generator $\mathcal{A}_\theta$ of (X_t) has a spectral gap $\lambda > 0$ under Condition 2.2.2. This leads to a well-known bound for the transition operator which we formulate as a separate lemma. In the following,

$$\mathscr{L}_0^2(\mu_\theta) = \left\{f : S \to \mathbb{R} : \mu_\theta(f^2) < \infty, \mu_\theta(f) = 0\right\}.$$

Lemma 2.3.2. *Let $f \in \mathscr{L}_0^2(\mu_\theta)$. Then,*

$$\left\| P_t^\theta f \right\|_2 \leq e^{-\lambda t} \|f\|_2 \tag{2.14}$$

for all $t \geq 0$.

As a consequence, $\|U_\theta(f)\|_2 < \infty$ for any $f \in \mathscr{L}_0^2(\mu_\theta)$ and we see that the potential operator

$$U_\theta : \mathscr{L}_0^2(\mu_\theta) \to \mathscr{L}^2(\mu_\theta)$$

is well-defined. It follows immediately from the defining property (2.13) that U_θ is linear in the sense that

$$U_\theta(\alpha f + \beta g) = \alpha U_\theta(f) + \beta U_\theta(g)$$

for scalars $\alpha, \beta \in \mathbb{R}$ and $f, g \in \mathscr{L}_0^2(\mu_\theta)$.

In this paper we generally restrict ourselves to the set of functions

$$\mathscr{H}_\theta^2 = \left\{ f \in \mathcal{C}_p^2(S) : \mu_\theta(f) = 0, U_\theta(f) \in \mathcal{C}_p^2(S) \right\}, \tag{2.15}$$

which ensures that $\mathcal{L}_\theta(U_\theta(f))$ is well-defined, $\mathcal{A}_\theta(U_\theta(f)) = \mathcal{L}_\theta(U_\theta(f))$ and $\mathscr{H}_\theta^2 \subset \mathscr{L}_0^2(\mu_\theta)$. The following result characterizes the potential $U_\theta(f)$ as the solution of the so-called *Poisson equation* for any $f \in \mathscr{H}_\theta^2$.

Proposition 2.3.3. *Let $f \in \mathscr{H}_\theta^2$. Then, $U_\theta(f)$ is a solution of the Poisson equation, i.e.*

$$\mathcal{L}_\theta(U_\theta(f)) = -f,$$

where $\mathcal{L}_\theta$ denotes the differential operator (2.9) *corresponding to the generator of (X_t).*

For our purposes, the existence of $U_\theta(f)$ for $f \in \mathscr{L}_0^2(\mu_\theta)$ is ensured by the ρ-mixing assumption on (X_t). General results on existence and regularity implications of the potential $U_\theta(f)$ for diffusion processes (X_t) and $f : S \to \mathbb{R}$ can be found in Pardoux and Veretennikov (2001).

2.3.3 Central limit theory

With Proposition 2.3.3 in place, we obtain the following CLT for $V_n(f)$. The proof requires the additional rate assumption that $n\Delta_n^3 \to 0$. Consistent with the general notation, we write $U_0 = U_{\theta_0}$, $\mathscr{H}_0^2 = \mathscr{H}_{\theta_0}^2$, etc., for the true parameter θ_0.

Proposition 2.3.4. *Let $f \in \mathscr{H}_0^2$. If $n\Delta_n^3 \to 0$, then*

$$\sqrt{n\Delta_n} V_n(f) = \sqrt{n\Delta_n} \left(\frac{1}{n} \sum_{i=1}^n f(X_{t_{i-1}^n}) \right) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mathcal{V}_0(f)\right),$$

where

$$\mathcal{V}_0(f) = \mu_0\left([\partial_x U_0(f) b(\,\cdot\,;\theta_0)]^2 \right) = 2\mu_0\left(f U_0(f) \right). \tag{2.16}$$

Remark 2.3.5. Compared to the low-frequency sampling scenario where $\Delta_n = \Delta > 0$, the integral construction of $U_\theta(f)$ in (2.13) can be interpreted as the limit as $\Delta \to 0$ of the discrete-time potential,

$$\tilde{U}_\theta(f) = \Delta \sum_{k=0}^\infty P_{k\Delta}^\theta f,$$

and the role of $U_0(f)$ in Proposition 2.3.4 is similar to that of $\tilde{U}_\theta(f)$ in the classic central limit theorem for functionals $\frac{1}{n}\sum_{i=1}^n f(X_{(i-1)\Delta})$; see e.g. Theorem 1, Florens-Zmirou (1989).

2.4 Asymptotic theory

In this section we present our main asymptotic results for prediction-based estimators. The proofs are based on general asymptotic theory for estimating functions in M. Sørensen (2012). For the most part, we restrict the discussion to estimating functions of the form (2.4) with $N = 1$ and, for simplicity, write

$$G_n(\theta) = \sum_{i=q}^{n} \pi_{i-1} \left[f(X_{t_i^n}) - \breve{\pi}_{i-1}(\theta) \right], \tag{2.17}$$

$\mathcal{P}_{i-1}$ for the corresponding predictor spaces, etc. The extension to multiple predictor functions $\{f_j\}_{j=1}^N$ is considered in Section 2.4.3.

2.4.1 Simple predictor spaces

The simplest class of estimating functions of the form (2.17) is obtained for $q = 0$, in which case $\mathcal{P}_{i-1} = \operatorname{span}\{1\}$. In this case, the orthogonal projection $\breve{\pi}_{i-1}(\theta) = \mu_\theta(f)$ and the one-dimensional predictor space $\mathcal{P}_{i-1}$ enables us to estimate a parameter $\theta \in \Theta \subset \mathbb{R}$. Consistently, we suppose that $d = 1$ in the following and study the one-dimensional estimating function

$$G_n(\theta) = \sum_{i=1}^{n} \left[f(X_{t_i^n}) - \mu_\theta(f) \right]. \tag{2.18}$$

For the asymptotic theory to carry through, we impose the following regularity conditions on G_n.

Condition 2.4.1. *Suppose that*

- $f \in \mathcal{C}_p^2(S)$,
- $f^*(x) := f(x) - \mu_0(f) \in \mathscr{H}_0^2$,
- $\theta \mapsto \mu_\theta(f) \in \mathcal{C}^1$.

By applying the limit theorems provided in Section 2.3, we easily identify conditions that ensure consistency and asymptotic normality of G_n-estimators for estimating functions that satisfy Condition 2.4.1.

Theorem 2.4.2. *Assume Condition 2.4.1 and suppose that $\partial_\theta \mu_\theta(f) \neq 0$ and that the identifiability condition*

$$(\mu_0 - \mu_\theta)(f) \neq 0,$$

holds for all $\theta \neq \theta_0$.

- *There exists a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ which, as $n \to \infty$, is unique in any compact subset $\mathcal{K} \subset \Theta$ containing θ_0 with $\mathbb{P}_0$-probability approaching one.*
- *If, moreover, $n\Delta_n^3 \to 0$, then*

$$\sqrt{n\Delta_n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, [\partial_\theta \mu_0(f)]^{-2}\, \mathcal{V}_0(f)\right), \tag{2.19}$$

where

$$\mathcal{V}_0(f) = 2\int_S f^*(x) U_0(f^*)(x) \mu_0(dx).$$

2.4.2 1-lag predictor spaces

The inclusion of past observations into the predictor space $\mathcal{P}_{i-1}$ raises the mathematical complexity dramatically. Our main result in this section (Theorem 2.4.5) shows that for $q = 1$, prediction-based G_n-estimators remain consistent and asymptotically normal under suitable regularity conditions.

For $q = 1$, the basis vector $Z_{i-1} = (1, f(X_{t_{i-1}^n}))^T$ and the normal equations (2.8) take the form

$$\mathbb{E}_\theta\left(\begin{pmatrix} 1 \\ f(X_0) \end{pmatrix} f(X_{\Delta_n})\right) - \mathbb{E}_\theta \begin{pmatrix} 1 & f(X_0) \\ f(X_0) & f^2(X_0) \end{pmatrix} \begin{pmatrix} \breve{a}_n(\theta)_0 \\ \breve{a}_n(\theta)_1 \end{pmatrix} = 0.$$

As a consequence,

$$\breve{\pi}_{i-1}(\theta) = \breve{a}_n(\theta)_0 + \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}),$$

where $\breve{a}_n(\theta)_0$ and $\breve{a}_n(\theta)_1$ are uniquely determined by the moment conditions

$$\begin{aligned} \breve{a}_n(\theta)_0 &= \mu_\theta(f)\,(1 - \breve{a}_n(\theta)_1)\,, \\ \breve{a}_n(\theta)_1 &= \frac{\mathbb{E}_\theta\left[f(X_0)f(X_{\Delta_n})\right] - \left[\mu_\theta(f)\right]^2}{\mathbb{V}ar_\theta f(X_0)}, \end{aligned}$$

and consistent with a two-dimensional predictor space $\mathcal{P}_{i-1}$, we suppose that $d = 2$ in the following and study the estimating function

$$G_n(\theta) = \sum_{i=1}^n \begin{pmatrix} 1 \\ f(X_{t_{i-1}^n}) \end{pmatrix} \left[f(X_{t_i^n}) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n})\right]. \tag{2.20}$$

As part of the proof of Lemma 2.4.4 below, we show that the projection coefficient $\breve{a}_n(\theta)$ has a power expansion

$$\breve{a}_n(\theta) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Delta_n \begin{pmatrix} -K_f(\theta)\mu_\theta(f) \\ K_f(\theta) \end{pmatrix} + \Delta_n^2 R(\Delta_n; \theta), \tag{2.21}$$

where $|R(\Delta_n;\theta)| \leq C(\theta)$ and the constant $K_f(\theta)$ is explicitly given by

$$K_f(\theta) = \frac{\mu_\theta(f\mathcal{L}_\theta f)}{\mathbb{V}ar_\theta f(X_0)}. \tag{2.22}$$

This observation enables us to formulate a set of regularity conditions on G_n for the asymptotic theory:

Condition 2.4.3. *Suppose that*

- $f \in \mathcal{C}_p^4(S)$,
- $f_1^*(x) = K_f(\theta_0)\left[\mu_0(f) - f(x)\right] \in \mathscr{H}_0^2$,
- $f_2^*(x) = f(x)\mathcal{L}_0 f(x) - K_f(\theta_0)f(x)\left[f(x) - \mu_0(f)\right] \in \mathscr{H}_0^2$,
- $\theta \mapsto \mu_\theta(f) \in \mathcal{C}^1$, $\theta \mapsto K_f(\theta) \in \mathcal{C}^1$ *and*

$$\sup_{\theta\in\mathcal{M}} \|\partial_{\theta^T} R(\Delta_n;\theta)\| \leq C(\mathcal{M}), \tag{2.23}$$

for a compact, convex subspace $\mathcal{M}$ containing θ_0 and Δ_n sufficiently small.

The matrix norm $\|\cdot\|$ in (2.23) can be chosen arbitrarily and we suppose for convenience that $\|\cdot\|$ is submultiplicative. The following lemma essentially implies the existence of a consistent sequence of G_n-estimators in Theorem 2.4.5. As the proof is somewhat long, we formulate it as a separate result.

Lemma 2.4.4. *Assume that Condition 2.4.3 holds. Then, for any $\theta \in \Theta$,*

$$(n\Delta_n)^{-1}G_n(\theta) \xrightarrow{\mathbb{P}_0} \gamma(\theta_0;\theta) = \begin{pmatrix} K_f(\theta)(\mu_\theta - \mu_0)(f) \\ \mu_0(f\mathcal{L}_0 f) - K_f(\theta)\left[\mu_0(f^2) - \mu_0(f)\mu_\theta(f)\right] \end{pmatrix}$$

and, moreover,

$$\sup_{\theta\in\mathcal{M}} \left\|(n\Delta_n)^{-1}\partial_{\theta^T} G_n(\theta) - W(\theta)\right\| \xrightarrow{\mathbb{P}_0} 0 \tag{2.24}$$

where

$$W(\theta) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix} \begin{pmatrix} \partial_{\theta_1}\left[K_f(\theta)\mu_\theta(f)\right] & \partial_{\theta_2}\left[K_f(\theta)\mu_\theta(f)\right] \\ -\partial_{\theta_1}K_f(\theta) & -\partial_{\theta_2}K_f(\theta) \end{pmatrix}.$$

Theorem 2.4.5. *Assume Condition 2.4.3 and suppose that $W(\theta)$ is non-singular and that the identifiability condition*

$$\gamma(\theta_0;\theta) \neq 0,$$

holds for all $\theta \neq \theta_0$.

- *There exists a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ which, as $n \to \infty$, is unique in any compact subset $\mathcal{K} \subset \Theta$ containing θ_0 with $\mathbb{P}_0$-probability approaching one.*
- *If, moreover, $n\Delta_n^3 \to 0$, then*

$$\sqrt{n\Delta_n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2\left(0, \left[W(\theta_0)^{-1}\mathcal{V}_0(f)(W(\theta_0)^{-1})^T\right]\right), \tag{2.25}$$

where

$$\mathcal{V}_0(f) = \begin{pmatrix} \mu_0\left([\partial_x U_0(f_1^*)b(\,\cdot\,;\theta_0)]^2\right) & \mathbb{C}ov(f) \\ \mathbb{C}ov(f) & \mu_0\left([\partial_x U_0(f_2^*) + f\partial_x f]^2\, b^2(\,\cdot\,;\theta_0)\right) \end{pmatrix},$$

with

$$\mathbb{C}ov(f) = \mu_0\left(\partial_x U_0(f_1^*)\left[\partial_x U_0(f_2^*) + f\partial_x f\right] b^2(\,\cdot\,;\theta_0)\right).$$

Remark 2.4.6. If we denote the estimating function (2.20) as

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

the proof of Lemma 2.4.4 shows that

$$\mathbb{E}_\theta\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,\middle|\,\mathcal{F}_{i-1}^n\right) = \Delta_n g^*(X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)$$

for a non-zero function g^* and $\theta \in \Theta$. Therefore, the estimating functions in this section lie outside the class of approximate martingale estimating functions defined in M. Sørensen (2017). In particular, the proof of asymptotic normality in Theorem 2.4.5 requires a bit of work since the remainder term obtained by compensating G_n is non-negligible.

2.4.3 Multiple predictor functions and optimal estimation

Estimating functions with multiple predictor functions,

$$G_n(\theta) = \sum_{i=q}^{n}\sum_{j=1}^{N} \pi_{i-1,j}\left[f_j(X_{t_i^n}) - \breve{\pi}_{i-1,j}(\theta)\right] \tag{2.26}$$

appear frequently in practice; see Section 2.6. In the following, we indicate how to extend the asymptotic theory from estimating functions with a single predictor function (2.17) to the more general case (2.26) and briefly consider optimal estimation in relation to over-identification of the parameter $\theta \in \Theta \subset \mathbb{R}^d$.

As emphasized in M. Sørensen (2011), the estimating functions (2.26) have a more compact vector representation

$$G_n(\theta) = A_n(\theta) \sum_{i=q}^{n} Z_{i-1} \left[F(X_{t_i^n}) - \breve{\Pi}_{i-1}(\theta) \right], \tag{2.27}$$

where $F(x) = (f_1(x), \ldots, f_N(x))^T$, $\breve{\Pi}_{i-1}(\theta) = (\breve{\pi}_{i-1,1}(\theta), \ldots, \breve{\pi}_{i-1,N}(\theta))^T$ and

$$Z_{i-1} = \begin{pmatrix} Z_{i-1,1} & 0_{q_1+1} & \cdots & 0_{q_1+1} \\ 0_{q_2+1} & Z_{i-1,2} & \cdots & 0_{q_2+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{q_N+1} & 0_{q_N+1} & \cdots & Z_{i-1,N} \end{pmatrix}. \tag{2.28}$$

Recall that $Z_{i-1,j}$ denotes the column vector (2.7) of basis elements of $\mathcal{P}_{i-1,j}$ and the notation 0_{q_j+1} denotes a column vector of length $q_j + 1$ containing zeroes only. Consistently, the dimension of Z_{i-1} in (2.28) is $\bar{d} \times N$ where $\bar{d} := N + \sum_{j=1}^{N} q_j$. The coefficient matrix $A_n(\theta)$ is $d \times \bar{d}$ to match a d-dimensional parameter θ.

A strong property of the vector representation (2.27) is that it enables us to extend the proofs in Section 2.9 from estimating functions with a single predictor function (2.17) to estimating functions of the more general form (2.26) by imposing that $A_n(\theta) \to A(\theta)$ as $n \to \infty$ and examining the normalized sum

$$V_n \times \sum_{i=q}^{n} Z_{i-1} \left[F(X_{t_i^n}) - \breve{\Pi}_{i-1}(\theta) \right],$$

where V_n is a diagonal $\bar{d} \times \bar{d}$ matrix,

$$V_n = \operatorname{diag}\left(v_{n,1}^{(1)}, \ldots, v_{n,q_1+1}^{(1)}, \ldots, v_{n,1}^{(N)}, \ldots, v_{n,q_N+1}^{(N)} \right),$$

and $v_{n,k_j}^{(j)} \to 0$ are appropriate normalization rates, e.g. $v_{n,k_j}^{(j)} = n^{-1}$ or $v_{n,k_j}^{(j)} = (n\Delta_n)^{-1}$.

The regularity condition $d \leq \bar{d}$ is necessary for θ to be identified by the estimating equation $G_n(\theta) = 0$ and we say that θ is *over-identified* if $d < \bar{d}$. Whereas Z_{i-1}, F and $\breve{\Pi}_{i-1}(\theta)$ are fully determined by our choice of predictor functions $\{f_j\}_{j=1}^N$ and corresponding predictor spaces $\{\mathcal{P}_{i-1,j}\}_j$, the coefficient matrix $A_n(\theta)$ can be chosen optimally if $d < \bar{d}$. The Godambe optimal estimating function G_n^* within the class (2.27) is the element for which the asymptotic variance of $\hat{\theta}_n$ in the error distribution

$$\sqrt{n\Delta_n} \left(\hat{\theta}_n - \theta_0 \right) \xrightarrow{\mathscr{D}_0} \mathcal{N}_d \left(0, \operatorname{AVAR}(\hat{\theta}_n) \right)$$

is smallest; see e.g. Godambe and Heyde (1987) or the monograph Heyde (1997). For our purposes, the optimality discussion in Section 3 in M. Sørensen (2011) covers the class of estimating functions (2.27) considered in this paper and Proposition 3.2 in M. Sørensen (2011) shows how to evaluate the optimal coefficient matrix $A_n^*(\theta)$.

2.5 Estimating the asymptotic variance

Estimation of the asymptotic variance (AVAR) of $\hat{\theta}_n$ is necessary for the construction of confidence intervals in practice. In this section we construct a Monte Carlo-based estimator for the class of estimating functions

$$G_n(\theta) = \sum_{i=1}^{n} \left[f(X_{t_i^n}) - \mu_\theta(f) \right] \tag{2.29}$$

defined in Section 2.4.1 and propose a feasible extension for the predictor spaces studied in Section 2.4.2. Moreover, we derive an upper bound for $\mathrm{AVAR}(\hat{\theta}_n)$ for estimating functions (2.29) by invoking the mixing property of (X_t) and show that it is tight by estimating the mean of an Ornstein-Uhlenbeck process.

2.5.1 Simple predictor spaces

For the construction we suppose that T is a random variable defined on an auxiliary probability space $(\Omega', \mathcal{F}', \mathbb{P}'_\gamma)$ such that $T \sim \exp(\gamma)$ and consider the product extension

$$\tilde{\Omega} = \Omega \times \Omega', \quad \tilde{\mathcal{F}} = \mathcal{F} \otimes \mathcal{F}', \quad \tilde{\mathbb{P}}_{\theta,\gamma} = \mathbb{P}_\theta \times \mathbb{P}'_\gamma.$$

In particular, this implies that

$$\tilde{\mathbb{E}}_{\theta,\gamma} f(X_0) = \int_{\tilde{\Omega}} f(X_0)(\omega) \tilde{\mathbb{P}}_{\theta,\gamma}(d\tilde{\omega}) = \int_{\Omega} f(X_0)(\omega) \mathbb{P}_\theta(d\omega) = \mathbb{E}_\theta f(X_0)$$

for any $f \in \mathscr{L}^1(\mu_\theta)$ and, similarly, we freely interchange expectation w.r.t. $\tilde{\mathbb{P}}_{\theta,\gamma}$ and $\mathbb{P}'_\gamma$ whenever they coincide in the following.

As shown in Theorem 2.4.2,

$$\mathrm{AVAR}(\hat{\theta}_n) = \frac{2\mu_0 \left(f^* U_0(f^*) \right)}{[\partial_\theta \mu_0(f)]^2} \tag{2.30}$$

for $f^* = f - \mu_0(f)$. To construct an explicit estimator, we note that for any $g \in \mathscr{H}_\theta^2$,

$$\begin{aligned}
\mu_\theta\left(gU_\theta(g)\right) &= \int_S g(x)\left(\int_0^\infty P_t^\theta g(x)\,\mathrm{d}t\right)\mu_\theta(dx)\\
&= \int_S\left(\int_0^\infty g(x)P_t^\theta g(x)\,\mathrm{d}t\right)\mu_\theta(dx)\\
&= \int_0^\infty\left(\int_S g(x)P_t^\theta g(x)\mu_\theta(dx)\right)\mathrm{d}t\\
&= \int_0^\infty\left(\int_S g(x)\tilde{\mathbb{E}}_{\theta,\gamma}\left(g(X_t)\mid X_0=x\right)\mu_\theta(dx)\right)\mathrm{d}t\\
&= \int_0^\infty\left(\int_S \tilde{\mathbb{E}}_{\theta,\gamma}\left(g(X_0)g(X_t)\mid X_0=x\right)\mu_\theta(dx)\right)\mathrm{d}t\\
&= \int_0^\infty \tilde{\mathbb{E}}_{\theta,\gamma}\left(g(X_0)g(X_t)\right)\mathrm{d}t,
\end{aligned}$$

where we apply Fubini's theorem to interchange the order of integration. Moreover, since the canonical extensions of (X_t) and T to variables on $\tilde{\Omega}$ are independent under the product measure $\tilde{\mathbb{P}}_{\theta,\gamma}$, it holds that

$$\begin{aligned}
\mu_\theta\left(gU_\theta(g)\right) &= \int_0^\infty \tilde{\mathbb{E}}_{\theta,\gamma}\left(g(X_0)g(X_t)\right)\mathrm{d}t\\
&= \gamma^{-1}\int_0^\infty e^{\gamma t}\tilde{\mathbb{E}}_{\theta,\gamma}\left(g(X_0)g(X_t)\right)\gamma e^{-\gamma t}\,\mathrm{d}t\\
&= \gamma^{-1}\int_0^\infty \tilde{\mathbb{E}}_{\theta,\gamma}\left(e^{\gamma T}g(X_0)g(X_T)\;\middle|\;T=t\right)\gamma e^{-\gamma t}\,\mathrm{d}t\\
&= \gamma^{-1}\tilde{\mathbb{E}}_{\theta,\gamma}\left[\tilde{\mathbb{E}}_{\theta,\gamma}\left(e^{\gamma T}g(X_0)g(X_T)\;\middle|\;T\right)\right]\\
&= \gamma^{-1}\tilde{\mathbb{E}}_{\theta,\gamma}\left[e^{\gamma T}g(X_0)g(X_T)\right].
\end{aligned}$$

As a consequence, if $\{T_i\}_{i=1}^K$ denote independent $\exp(\gamma)$ variables under $\mathbb{P}'_\gamma$ and $\{\tilde{X}_{T_i}^{(i)}\}$ the values of independent trajectories of (X_t) under $\mathbb{P}_\theta$ with initial value $X_0 = x$, the estimator

$$\gamma^{-1}\frac{1}{K}\sum_{i=1}^K e^{\gamma T_i} g\left(\tilde{X}_0^{(i)}\right) g\left(\tilde{X}_{T_i}^{(i)}\right) \tag{2.31}$$

converges $\tilde{\mathbb{P}}_{\theta,\gamma}$-almost surely towards $\mu_\theta\left(gU_\theta(g)\right)$ as $K \to \infty$ for any $g \in \mathscr{H}_\theta^2$.

Based on this observation, the following algorithm can be used in practice to estimate $\mathrm{AVAR}(\hat{\theta}_n)$:

Monte Carlo Estimation

1. *Determine* $\hat{\theta}_n$*,*
2. *Simulate* K *independent variables* $T_i \sim \exp(\gamma)$ *for a fixed* $\gamma > 0$*,*
3. *Simulate* $t \mapsto \tilde{X}_t^{(i)}$ *on* $[0, T_i]$ *under* $\mathbb{P}_{\hat{\theta}_n}$*,*
4. *Evaluate*

$$\widehat{\mathrm{AVAR}}(\hat{\theta}_n) = 2 \cdot [\partial_\theta \mu_{\hat{\theta}_n}(f)]^{-2} \gamma^{-1} \frac{1}{K} \sum_{i=1}^{K} e^{\gamma T_i} \hat{f}^* \left(\tilde{X}_0^{(i)}\right) \hat{f}^* \left(\tilde{X}_{T_i}^{(i)}\right), \tag{2.32}$$

where $\hat{f}^*(x) := f(x) - \mu_{\hat{\theta}_n}(f)$*.*

Remark 2.5.1. If the transition density of (X_t) is known, the use of an exact simulation scheme reduces step 3 to sampling $\tilde{X}_0^{(i)}$ under $\mathbb{P}_{\hat{\theta}_n}$ and, subsequently, $\tilde{X}_{T_i}^{(i)}$. Alternatively, an appropriate discretization scheme is necessary to simulate the value of $\tilde{X}_{T_i}^{(i)}$.

In addition, the mixing property of (X_t) leads to the following upper bound for $\mathrm{AVAR}(\hat{\theta}_n)$.

Proposition 2.5.2. *Suppose that* (X_t) *and* $G_n(\theta)$ *satisfy Condition 2.2.2 and Condition 2.4.1, respectively, and let* λ_0 *denote the spectral gap of* (X_t) *under* $\mathbb{P}_0$*. Then,*

$$\mathrm{AVAR}(\hat{\theta}_n) \leq \left(\frac{2}{\lambda_0}\right) \frac{\mathbb{V}ar_0 f(X_0)}{[\partial_\theta \mu_0(f)]^2}. \tag{2.33}$$

Example 2.5.3. The Ornstein-Uhlenbeck process

$$dX_t = \kappa(\eta - X_t)dt + \xi dB_t,$$

is a well-known diffusion model that satisfies Condition 2.2.2. It was introduced in finance by Vasicek (1977) to model interest rate dynamics. In particular, the invariant distribution of (X_t) under $\mathbb{P}_\theta$ is $\mu_\theta \sim \mathcal{N}\left(\eta, \frac{\xi^2}{2\kappa}\right)$.

If we suppose that the values of κ and ξ are known, estimation of η provides an illustrative example where we can determine $\mathrm{AVAR}(\hat{\theta}_n)$ explicitly and attain the (feasible) upper bound of Proposition 2.5.2. To identify η, we choose $f(x) = x$ and by direct calculation,

$$
\begin{aligned}
U_0(f^*)(x) &= \int_0^\infty \mathbb{E}_0\left(f^*(X_t) \mid X_0 = x\right) \mathrm{d}t \\
&= \int_0^\infty \left[\mathbb{E}_0\left(X_t \mid X_0 = x\right) - \eta_0\right] \mathrm{d}t \\
&= \int_0^\infty \left[xe^{-\kappa t} + \eta_0\left(1 - e^{-\kappa t}\right) - \eta_0\right] \mathrm{d}t \\
&= (x - \eta_0) \int_0^\infty e^{-\kappa t} \,\mathrm{d}t \\
&= \frac{(x - \eta_0)}{\kappa}.
\end{aligned}
$$

As a consequence,

$$
\mu_0\left(f^* U_0(f^*)\right) = \frac{1}{\kappa} \int_{\mathbb{R}} (x - \eta_0)^2 \mu_0(dx) = \frac{\xi^2}{2\kappa^2} \tag{2.34}
$$

and, in turn,

$$
\mathrm{AVAR}(\hat{\theta}_n) = \left(\frac{\xi}{\kappa}\right)^2.
$$

That we attain the bound (2.33) follows by observing that $\partial_\eta \mathbb{E}_0(X_0) = 1$, $\mathbb{V}ar_0(X_0) = \frac{\xi^2}{2\kappa}$ and applying that the spectral gap $\lambda_0 = \kappa$.

2.5.2 1-lag predictor spaces

For the prediction-based estimating functions in Section 2.4.2, Theorem 2.4.5 shows that the asymptotic variance

$$
\mathrm{AVAR}(\hat{\theta}_n) = W(\theta_0)^{-1} \mathcal{V}_0(f) (W(\theta_0)^{-1})^T, \tag{2.35}
$$

where

$$
W(\theta) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix} \begin{pmatrix} \partial_{\theta_1}\left[K_f(\theta)\mu_\theta(f)\right] & \partial_{\theta_2}\left[K_f(\theta)\mu_\theta(f)\right] \\ -\partial_{\theta_1} K_f(\theta) & -\partial_{\theta_2} K_f(\theta) \end{pmatrix},
$$

and

$$
\mathcal{V}_0(f) = \begin{pmatrix} \mu_0\left(\left[\partial_x U_0(f_1^*) b(\,\cdot\,;\theta_0)\right]^2\right) & \mu_0\left(\partial_x U_0(f_1^*)\left[\partial_x U_0(f_2^*) + f\partial_x f\right] b^2(\,\cdot\,;\theta_0)\right) \\ -\text{''}- & \mu_0\left(\left[\partial_x U_0(f_2^*) + f\partial_x f\right]^2 b^2(\,\cdot\,;\theta_0)\right) \end{pmatrix}.
$$

To construct an estimator in this case, two important issues need be addressed. Firstly, since

$$
K_f(\theta) = \frac{\mu_\theta(f\mathcal{L}_\theta f)}{\mathbb{V}ar_\theta f(X_0)},
$$

the partial derivatives $\partial_{\theta_j}[K_f(\hat{\theta}_n)\mu_{\hat{\theta}_n}(f)]$ and $\partial_{\theta_j} K_f(\hat{\theta}_n)$ may be approximated using h-step difference quotients for a fixed $h > 0$. Note, however, that if the evaluation of K_f is based on simulation, it is crucial to apply the same pseudo-random sequence from $\mu_{\hat{\theta}_n}$ to avoid numerical instability.

Secondly, although f_1^* and f_2^* are easily approximated by replacing θ_0 with $\hat{\theta}_n$ as in (2.32), only

$$\mu_0\left([\partial_x U_0(f_1^*)b(\,\cdot\,;\theta_0)]^2\right) = \mu_0\left(f_1^* U_0(f_1^*)\right) \tag{2.36}$$

can be approximated using an estimator of the type (2.31). The remaining entries of $\mathcal{V}_0(f)$ contain integrals involving $\partial_x U_0(f_1^*)$ and $\partial_x U_0(f_2^*)$. Reasoning as in Section 2.5.1, our solution will be to construct a pointwise estimator of $\partial_x U_\theta(g)(x)$ for general $g \in \mathscr{H}_\theta^2$ using the difference quotient as an approximation. This, in turn, enables us to approximate the integrals w.r.t μ_0 by sampling from $\mu_{\hat{\theta}_n}$. Specifically, for $h > 0$ small,

$$\partial_x U_\theta(g)(x) \approx \frac{U_\theta(g)(x+h) - U_\theta(g)(x)}{h}$$

and re-applying the product space $(\tilde{\Omega}, \tilde{\mathcal{F}}, \tilde{\mathbb{P}}_{\theta,\gamma})$ defined in Section 2.5.1,

$$\begin{aligned}
U_\theta(g)(y) &= \int_0^\infty P_t^\theta g(y)\,\mathrm{d}t \\
&= \gamma^{-1}\int_0^\infty e^{\gamma t}\tilde{\mathbb{E}}_{\theta,\gamma}\left(g(X_t) \mid X_0 = y\right)\gamma e^{-\gamma t}\,\mathrm{d}t \\
&= \gamma^{-1}\int_0^\infty \tilde{\mathbb{E}}_{\theta,\gamma}\left(e^{\gamma t} g(X_t) \,\middle|\, X_0 = y\right)\gamma e^{-\gamma t}\,\mathrm{d}t \\
&= \gamma^{-1}\tilde{\mathbb{E}}_{\theta,\gamma}\left(e^{\gamma T} g(X_T) \,\middle|\, X_0 = y\right) \\
&\approx \gamma^{-1}\frac{1}{K}\sum_{i=1}^K e^{\gamma T_i} g\left(\tilde{X}_{T_i}^{(i)} \,\middle|\, \tilde{X}_0^{(i)} = y\right),
\end{aligned}$$

where we explicitly emphasize the dependence on y. Hence, if we sample independent variables $\{T_i\}_{i=1}^K$ and apply a fixed seed for the evaluation of $\{\tilde{X}_{T_i}^{(i)}\}$, we obtain the pointwise estimator

$$\widehat{\partial_x U}_\theta(g)(x) = (h\gamma)^{-1}\frac{1}{K}\sum_{i=1}^K e^{\gamma T_i}\left[g\left(\tilde{X}_{T_i}^{(i)} \,\middle|\, \tilde{X}_0^{(i)} = x+h\right) - g\left(\tilde{X}_{T_i}^{(i)} \,\middle|\, \tilde{X}_0^{(i)} = x\right)\right].$$

2.6 Applications in finance

To complement the asymptotic theory, we construct an explicit estimating function for the square-root (CIR) process and a simulation-based extension for the volatility process of a 3/2 model from finance. While explicit prediction-based estimating functions are generally available for the versatile class of Pearson diffusions defined in Forman and M. Sørensen (2008), the latter illustrates a simple way to encompass diffusion models for which analytic moment conditions are not readily available into the estimation framework.

Example 1 The square-root model

$$dX_t = \kappa(\eta - X_t)dt + \xi\sqrt{X_t}dB_t, \tag{2.37}$$

is widely used in many scientific areas. In finance, it was introduced as a model for interest rate dynamics by Cox, Ingersoll, and Ross (1985) and famously adopted by Heston (1993) to model the volatility of financial assets in relation to options pricing.

Under $\mathbb{P}_\theta$, the process has a unique invariant Gamma distribution with density

$$\mu_\theta(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1}e^{-\beta x}, \quad x \in \mathbb{R}_+ \tag{2.38}$$

for $\alpha = \frac{2\kappa\eta}{\xi^2}$ and $\beta = \frac{2\kappa}{\xi^2}$. In particular, if $X_0 \sim \mu_\theta$ and we impose the *Feller condition* $\alpha \geq 1$, the square-root process (2.37) satisfies Condition 2.2.2. A clear exposition of the fact that (X_t) is ρ-mixing under $\mathbb{P}_\theta$ can be found in Genon-Catalot, Jeantheau, and Larédo (2000).

To estimate the full parameter $\theta = (\kappa, \eta, \xi)$, we apply the pair of predictor spaces where $X_{t_i^n}$ is projected onto $\mathcal{P}_{i-1,1} = \text{span}\{1, X_{t_{i-1}^n}\}$, and $X_{t_i^n}^2$ onto $\mathcal{P}_{i-1,2} = \text{span}\{1\}$. As shown in Section 2.4, this yields the projections

$$\breve{\pi}_{i-1,1}(\theta) = \mathbb{E}_\theta(X_0)\left[1 - \breve{a}_n(\theta)_{11}\right] + \breve{a}_n(\theta)_{11}X_{t_{i-1}^n},$$

where

$$\breve{a}_n(\theta)_{11} = \frac{\mathbb{E}_\theta\left[X_0X_{\Delta_n}\right] - \mathbb{E}_\theta(X_0)^2}{\mathbb{V}ar_\theta X_0} \tag{2.39}$$

and $\breve{\pi}_{i-1,2}(\theta) = \breve{a}_n(\theta)_{20} = \mathbb{E}_\theta(X_0^2)$. The first and second moment of the invariant Gamma distribution are $\mathbb{E}_\theta(X_0) = \eta$ and $\mathbb{E}_\theta(X_0^2) = \eta\left(\eta + \frac{\xi^2}{2\kappa}\right)$, respectively, and using that

$$X_{\Delta_n} \mid X_0 \sim \frac{\xi^2(1 - e^{-\kappa\Delta_n})}{4\kappa} \cdot \chi^2_{2\alpha}\left(\frac{4\kappa e^{-\kappa\Delta_n}}{\xi^2(1 - e^{-\kappa\Delta_n})} \cdot X_0\right),$$

where $\chi_d^2(\lambda)$ denotes a non-central χ^2-distribution with d degrees of freedom and non-centrality parameter $\lambda > 0$, the conditional expectation

$$\mathbb{E}_\theta\left(X_{\Delta_n} \mid \mathcal{F}_0\right) = X_0e^{-\kappa\Delta_n} + \eta\left(1 - e^{-\kappa\Delta_n}\right),$$

and, hence, by the tower property,

$$\mathbb{E}_\theta(X_0X_{\Delta_n}) = \eta\left(\eta + \frac{\xi^2}{2\kappa}\right)e^{-\kappa\Delta_n} + \eta^2\left(1 - e^{-\kappa\Delta_n}\right) = \eta^2 + \frac{\eta\xi^2}{2\kappa}e^{-\kappa\Delta_n}.$$

This shows that $\breve{a}_n(\theta)_{11} = e^{-\kappa\Delta_n}$ and we obtain the explicit estimating function

$$G_n(\theta) = \sum_{i=1}^n \begin{pmatrix} 1 & 0 \\ X_{t_{i-1}^n} & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} X_{t_i^n} - \eta\left(1 - e^{-\kappa\Delta_n}\right) - e^{-\kappa\Delta_n}X_{t_{i-1}^n} \\ X_{t_i^n}^2 - \eta\left(\eta + \frac{\xi^2}{2\kappa}\right) \end{pmatrix}. \tag{2.40}$$

Example 2 Motivated by the empirical findings in Aït-Sahalia (1996b), Ahn and Gao (1999) model interest rate dynamics using a diffusion process

$$dX_t = \kappa X_t(\eta - X_t)dt + \xi X_t^{3/2} dB_t, \tag{2.41}$$

where $\theta = (\kappa, \eta, \xi) \in \mathbb{R}_+^3$. Notably, this process differs from a square-root process by having a nonlinear drift term but remains a tractable choice for volatility dynamics; see e.g. Drimus (2012) for pricing and hedging of options under the so-called 3/2 model.

By Itô's formula, the inverse $V_t = 1/X_t$ is a square-root model

$$dV_t = (\kappa + \xi^2 - \kappa\eta V_t)dt - \xi\sqrt{V_t}dB_t,$$

from which we easily derive stationarity and mixing properties of (X_t); under $\mathbb{P}_\theta$, the process has an invariant inverse Gamma distribution with density

$$\mu_\theta(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\frac{\beta}{x}}, \quad x \in \mathbb{R}_+ \tag{2.42}$$

for $\alpha = \frac{2(\kappa+\xi^2)}{\xi^2}$ and $\beta = \frac{2\kappa\eta}{\xi^2}$. Moreover, since

$$\frac{2(\kappa + \xi^2)}{\xi^2} \geq 1 \Leftrightarrow 2\kappa + \xi^2 \geq 0,$$

(V_t) satisfies the Feller condition for any $\theta \in \mathbb{R}_+^3$. Hence, it follows from the construction of the mixing coefficient $\rho_X(t)$ in (2.10) that with $X_0 \sim \mu_\theta$, the diffusion model (2.41) is ρ-mixing as it generates the same σ-algebras as (V_t).

As a property of the inverse Gamma distribution, the ν'th moment of μ_θ is finite if and only if $\alpha > \nu$, which translates to the explicit parameter restriction

$$2\kappa + (2-\nu)\xi^2 > 0.$$

In particular, the second moment is always finite, but for $\nu \geq 3$ we must choose

$$\kappa > \frac{(\nu-2)}{2}\xi^2.$$

To construct an estimating function, we again consider the predictor spaces where we project $X_{t_i^n}$ onto $\mathcal{P}_{i-1,1} = \text{span}\{1, X_{t_{i-1}^n}\}$ and $X_{t_i^n}^2$ onto $\mathcal{P}_{i-1,2} = \text{span}\{1\}$. The first and second moments are $\mathbb{E}_\theta(X_0) = \frac{2\kappa\eta}{2\kappa+\xi^2}$ and $\mathbb{E}_\theta(X_0^2) = \frac{2\kappa\eta^2}{2\kappa+\xi^2}$, respectively, and by transforming the non-central χ^2-density of (V_t), one can show that

$$\mathbb{E}_\theta\left(X_{\Delta_n} \mid \mathcal{F}_0\right) = c \cdot q^{-1} \cdot e^{-u} \cdot M\left(q, q+1, u\right),$$

where $M(a,b,x)$, $b > a > 0$, denotes the confluent hypergeometric function

$$M(a,b,x) = \frac{\Gamma(b)}{\Gamma(b-a)\Gamma(a)} \int_0^1 e^{xy} y^{a-1} (1-y)^{b-a-1} \, \mathrm{d}y, \tag{2.43}$$

and

$$\begin{aligned} c(\theta) &= \frac{2\kappa\eta}{\xi^2\left(1-e^{-\kappa\eta\Delta_n}\right)}, \\ q(\theta) &= \frac{2(\kappa+\xi^2)}{\xi^2}-1, \\ u(X_0;\theta) &= X_0^{-1}\cdot c\cdot e^{-\kappa\eta\Delta_n}. \end{aligned}$$

By the tower property,

$$\mathbb{E}_\theta(X_0X_{\Delta_n}) = c\cdot q^{-1}\cdot\mathbb{E}_\theta\left(X_0\cdot e^{-u}\cdot M\left(q,q+1,u\right)\right). \tag{2.44}$$

The unconditional moment in (2.44) does not have a closed-form analytic representation in terms of θ, but can easily be approximated by simulating from the invariant distribution, i.e.

$$\mathbb{E}_\theta(X_0X_{\Delta_n}) \simeq c\cdot q^{-1}\cdot\frac{1}{K}\sum_{i=1}^{K}\tilde{X}_0^{(i)}\cdot e^{-\tilde{u}^{(i)}}\cdot M\left(q,q+1,\tilde{u}^{(i)}\right), \tag{2.45}$$

where $\{\tilde{X}_0^{(i)}\}_{i=1}^K$ are i.i.d. random variables drawn from (2.42). Simulated moment estimators have been studied by e.g. Duffie and Singleton (1993) to address this type of problems.

Hereby, we obtain the simulated estimating function

$$\tilde{G}_n(\theta) = \sum_{i=1}^{n}\begin{pmatrix} 1 & 0 \\ X_{t_{i-1}^n} & 0 \\ 0 & 1\end{pmatrix}\begin{pmatrix} X_{t_i^n}-\frac{2\kappa\eta}{2\kappa+\xi^2}\left(1-\tilde{a}_n(\theta)_{11}\right)-\tilde{a}_n(\theta)_{11}X_{t_{i-1}^n} \\ X_{t_i^n}^2-\frac{2\kappa\eta^2}{2\kappa+\xi^2}\end{pmatrix}, \tag{2.46}$$

where

$$\tilde{a}_n(\theta)_{11} = \frac{(2\kappa+\xi^2)^2}{2\kappa\eta^2\xi^2}\left[c\cdot q^{-1}\cdot\frac{1}{K}\sum_{i=1}^{K}\tilde{X}_0^{(i)}\cdot e^{-\tilde{u}^{(i)}}\cdot M\left(q,q+1,\tilde{u}^{(i)}\right)-\left(\frac{2\kappa\eta}{2\kappa+\xi^2}\right)^2\right],$$

and we obtain an estimator by solving $\tilde{G}_n(\theta)=0$.

2.7 Numerical results

Based on simulated trajectories, we evaluate the finite-sample behaviour of the estimating functions derived in Section 2.6. We do this for varying values of Δ_n while keeping the time horizon T fixed. Both processes are simulated exactly by sampling from the transition density of a square-root process; see e.g. Broadie and Kaya (2006).

Example 1 (cont.) By solving $G_n(\theta) = 0$ for the estimating function (2.40), we obtain the estimators

$$\begin{aligned}
\hat{\eta} &= \frac{1}{n}\sum_{i=0}^{n} X_{t_i^n} + \frac{e^{-\hat{\kappa}\Delta_n} X_{t_n^n} - X_0}{n\left(1 - e^{-\hat{\kappa}\Delta_n}\right)}, \\
0 &= \sum_{i=1}^{n} X_{t_{i-1}^n} X_{t_i^n} - \hat{\eta}\left(1 - e^{-\hat{\kappa}\Delta_n}\right)\sum_{i=1}^{n} X_{t_{i-1}^n} - e^{-\hat{\kappa}\Delta_n}\sum_{i=1}^{n} X_{t_{i-1}^n}^2, \\
\hat{\xi}^2 &= \frac{2\hat{\kappa}}{\hat{\eta}}\left[\left(\frac{1}{n}\sum_{i=1}^{n} X_{t_{i-1}^n}^2\right) - \hat{\eta}^2\right],
\end{aligned}$$

which are easily found numerically. As a measure of performance, we evaluate the empirical mean and standard deviation of our estimators based on $N = 50$, $N = 200$ and $N = 500$ trajectories of (X_t). The time horizon $T = 200$ is fixed. A summary of our simulation study is given in Table 2.1. With no exceptions, the estimator seems to approximate the true parameter $\theta_0 = (1, 10, 4)$ well.

TABLE 2.1

SUMMARY: SIMULATED SQUARE-ROOT PROCESS

N	Δ_n	$\hat{\kappa}$		$\hat{\eta}$		$\hat{\xi}$	
		mean	s.d.	mean	s.d.	mean	s.d.
50	1/2	1.07	0.20	9.85	0.76	3.97	0.26
50	1/4	1.05	0.18	10.03	0.90	3.98	0.13
50	0.1	1.03	0.16	9.97	0.82	3.99	0.10
200	1/2	1.07	0.21	10.03	0.90	3.97	0.22
200	1/4	1.05	0.17	10.02	0.89	3.99	0.14
200	0.1	1.04	0.15	9.98	0.89	4.00	0.10
500	1/2	1.06	0.20	10.00	0.91	4.00	0.24
500	1/4	1.05	0.17	9.97	0.87	3.99	0.15
500	0.1	1.04	0.16	10.02	0.90	4.00	0.10

Notes: (i) The standard deviation of $\hat{\eta}$ is relatively large. This is due to our choice of $\xi \gg \kappa$, which causes the trajectories $\{X_{t_i^n}\}_{i=0}^n$ to fluctuate substantially around the mean. For smaller values of ξ, the standard deviation decreases sharply, as expected.

Example 2 (cont.) For simulated estimating functions, we rely on numerical optimization in order to determine $\hat{\theta}_n$. The defining property $\tilde{G}_n(\hat{\theta}_n) = 0$ may equivalently be formulated as

$$\tilde{G}_n(\hat{\theta}_n)^T \tilde{G}_n(\hat{\theta}_n) = 0 \in \mathbb{R},$$

and in practice we determine $\hat{\theta}_n$ by solving the minimization problem

$$\hat{\theta}_n = \arg\min_{\theta \in \mathcal{K}} \tilde{G}_n(\theta)^T \tilde{G}_n(\theta) \tag{2.47}$$

on some bounded subset $\mathcal{K} \subset \Theta$. Note that for the implementation of $\tilde{G}_n(\theta)$ it is crucial to apply the same inverse Gamma sample (i.e. the same seed) for each value of θ; this ensures that the mapping $\theta \mapsto \tilde{G}_n(\theta)$ is sufficiently smooth for the optimization algorithm to converge. For our simulation study we choose $\theta_0 = (10, 7, 3)$. A typical trajectory of (X_t) is shown in Fig. 2.2 and our numerical results are summarized in Table 2.3.

Remark 2.7.1. To initialize the optimization algorithm reasonably close to the true parameter θ_0 we can, for small values of Δ_n, apply the empirical approximations

$$\begin{aligned}
\eta &= \frac{\mathbb{E}_\theta(X_0^2)}{\mathbb{E}_\theta(X_0)} \simeq \frac{\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}^n}^2}{\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}^n}}, \\
\xi^2 &= \frac{[\log X]_T}{\int_0^T X_s \, \mathrm{d}s} \simeq \frac{\sum_{i=1}^n \left(\log X_{t_i^n} - \log X_{t_{i-1}^n}\right)^2}{\Delta_n \sum_{i=1}^n X_{t_{i-1}^n}},
\end{aligned}$$

and subsequently solve $\mathbb{E}_\theta(X_0) = \frac{2\kappa\eta}{2\kappa + \xi^2}$ w.r.t. κ.

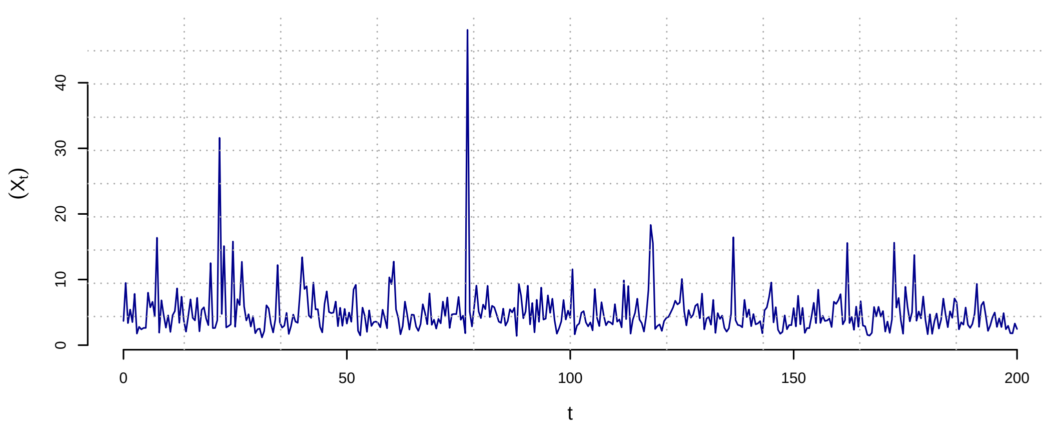

FIGURE 2.2: A typical realization of (X_t) for $\theta_0 = (10, 7, 3)$. In particular, the nonlinear drift term implies rapid mean-reversion when the process is above the stationary mean $\mathbb{E}_0(X_0) \approx 4.83$. Here $\Delta_n = 1/2$ and $T = 200$.

Table 2.3

Summary: simulated 3/2 diffusion

K	Δ_n	$\hat{\kappa}$	$\hat{\eta}$	$\hat{\xi}$
500	1/2	9.86	6.50	2.82
500	1/4	10.10	7.08	3.17
500	1/8	10.08	6.92	2.94
1000	1/2	10.02	6.42	2.45
1000	1/4	10.11	7.17	3.29
1000	1/8	10.24	7.77	3.93
2000	1/2	10.20	6.53	2.90
2000	1/4	10.24	7.12	3.24
2000	1/8	10.22	6.98	3.06

Notes: (i) The constant K determines the accuracy of the moment approximation; see (2.45). *(ii) The values of $\hat{\kappa}$, $\hat{\eta}$ and $\hat{\xi}$ are from particular realizations of (X_t) but suffice to illustrate convergence.*

2.8 Extensions and concluding remarks

An extension of interest would be to include observation error into the asymptotic theory, corresponding to microstructure noise; see e.g. Aït-Sahalia, Mykland, and Zhang (2005) and Gloter and Jacod (2001a,b) for results in this direction. For discretized diffusions, the use of finite-dimensional predictor spaces implies a loss of efficiency compared to the martingale approach in M. Sørensen (2017), however, for non-Markovian models, martingale methods are no longer tractable and an extension of the asymptotic theory to discretely observed integrated diffusion processes is provided in Chapter 3. Prediction-based estimating functions can also be applied for inference in diffusion-type stochastic volatility models, but requires the use of a different set of high-frequency limit theorems.

2.9 Appendix A: Proofs

Proof of Lemma 2.3.2. By Theorem 2.4 in Genon-Catalot, Jeantheau, and Larédo (2000), the mixing coefficient

$$\rho_X(t) = \sup\left\{\frac{\|P_t^\theta f\|_2}{\|f\|_2} : f \in \mathscr{L}_0^2(\mu_\theta)\right\},$$

and since (X_t) is reversible under Condition 2.2.2, Theorem 2.6 in Genon-Catalot, Jeantheau, and Larédo (2000) implies that $\rho_X(t) = e^{-\lambda t}$, where $\lambda > 0$ denotes the spectral gap of $\mathcal{A}_\theta$. In particular,

$$\frac{\left\|P_t^\theta f\right\|_2}{\|f\|_2} \leq e^{-\lambda t}$$

for any $f \in \mathscr{L}_0^2(\mu_\theta)$. $\square$

Proof of Proposition 2.3.3. Let $U_\theta^{(n)}(f) = \int_0^n P_t^\theta f \, \mathrm{d}t$. By Property P4 in Hansen and Scheinkman (1995), $U_\theta^{(n)}(f) \in \mathcal{D}_{\mathcal{A}_\theta}$ for each $n \in \mathbb{N}$ and

$$\lim_{n\to\infty} \mathcal{A}_\theta \left(U_\theta^{(n)}(f)\right) = \lim_{n\to\infty} \left[P_n^\theta f - f\right] = -f,$$

where limits are w.r.t. $\|\cdot\|_2$ and the latter equality holds since

$$\lim_{n\to\infty} \left\|P_n^\theta f\right\|_2 \leq \|f\|_2 \lim_{n\to\infty} e^{-\lambda n} = 0.$$

Moreover, by Jensen's inequality, Fubini's theorem and Lemma 2.3.2,

$$\begin{aligned}
\left\|U_\theta(f) - U_\theta^{(n)}(f)\right\|_2^2 &= \int_S \left(\int_n^\infty P_t^\theta f(x) \, \mathrm{d}t\right)^2 \mu_\theta(dx) \\
&= \int_S \left(\int_0^\infty 1\{t \geq n\} \lambda^{-1} e^{\lambda t} P_t^\theta f(x) \lambda e^{-\lambda t} \, \mathrm{d}t\right)^2 \mu_\theta(dx) \\
&\leq \int_S \left(\int_0^\infty 1\{t \geq n\} \lambda^{-2} e^{2\lambda t} \left(P_t^\theta f(x)\right)^2 \lambda e^{-\lambda t} \, \mathrm{d}t\right) \mu_\theta(dx) \\
&= \int_0^\infty \left(\int_S 1\{t \geq n\} \lambda^{-1} e^{\lambda t} \left(P_t^\theta f(x)\right)^2 \mu_\theta(dx)\right) \mathrm{d}t \\
&= \lambda^{-1} \int_n^\infty e^{\lambda t} \left(\int_S \left(P_t^\theta f(x)\right)^2 \mu_\theta(dx)\right) \mathrm{d}t \\
&= \lambda^{-1} \int_n^\infty e^{\lambda t} \cdot ||P_t^\theta f||_2^2 \, \mathrm{d}t \\
&\leq \lambda^{-1} \|f\|_2^2 \int_n^\infty e^{-\lambda t} \, \mathrm{d}t \\
&= \lambda^{-2} \|f\|_2^2 \, e^{-\lambda n},
\end{aligned}$$

which shows that $U_\theta^{(n)}(f)$ converges to $U_\theta(f)$ in $\mathscr{L}^2(\mu_\theta)$ as $n \to \infty$ and, taking $n = 0$,

$$\|U_\theta(f)\|_2 \leq \lambda^{-1} \|f\|_2 . \tag{2.48}$$

Using that $\mathcal{A}_\theta$ is closed and linear, we conclude that $\mathcal{A}_\theta \left(U_\theta(f)\right) = \mathcal{L}_\theta \left(U_\theta(f)\right) = -f$; see e.g. Property P7, Hansen and Scheinkman (1995). $\square$

Proof of Proposition 2.3.4. The proof is an application of the continuous-time central limit theorem for martingales. Firstly, we note that

$$\begin{aligned}
\frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} f(X_s)\,\mathrm{d}s &= \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} f(X_s)\,\mathrm{d}s \\
&= \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \left[f(X_s) - f(X_{t_{i-1}^n}) \right] \mathrm{d}s + \sqrt{n\Delta_n} V_n(f),
\end{aligned}$$

and our first step will be to show that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \left[f(X_s) - f(X_{t_{i-1}^n}) \right] \mathrm{d}s = o_{\mathbb{P}_0}(1). \tag{2.49}$$

If we let $A_i := \int_{(i-1)\Delta_n}^{i\Delta_n} \left[f(X_s) - f(X_{t_{i-1}^n}) \right] \mathrm{d}s$, Fubini's theorem combined with Lemma 2.10.2 implies that

$$\mathbb{E}_0 \left(A_i \mid \mathcal{F}_{i-1}^n \right) = \int_0^{\Delta_n} u \cdot R(u, X_{t_{i-1}^n}; \theta_0)\,\mathrm{d}u \leq_C \Delta_n^2 F(X_{t_{i-1}^n}; \theta_0)$$

for a generic function $F(x; \theta_0)$ of polynomial growth in x. Hence, under the additional rate assumption $n\Delta_n^3 \to 0$, it follows by Lemma 2.3.1 that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \mathbb{E}_0 \left(A_i \mid \mathcal{F}_{i-1}^n \right) \leq_C (n\Delta_n^3)^{1/2} \frac{1}{n} \sum_{i=1}^n F(X_{t_{i-1}^n}; \theta_0) \xrightarrow{\mathbb{P}_0} 0.$$

Moreover, for all $k \geq 1$, Jensen's inequality implies that

$$\begin{aligned}
|A_i|^k &= \Delta_n^k \left| \frac{1}{\Delta_n} \int_{(i-1)\Delta_n}^{i\Delta_n} \left[f(X_s) - f(X_{t_{i-1}^n}) \right] \mathrm{d}s \right|^k \\
&\leq \Delta_n^{k-1} \int_{(i-1)\Delta_n}^{i\Delta_n} |f(X_s) - f(X_{t_{i-1}^n})|^k\,\mathrm{d}s \\
&\leq \Delta_n^k \sup_{u \in [0, \Delta_n]} |f(X_{t_{i-1}^n + u}) - f(X_{t_{i-1}^n})|^k,
\end{aligned}$$

and, hence, by Lemma 2.10.1,

$$\begin{aligned}
\frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_0 \left(|A_i|^2 \mid \mathcal{F}_{i-1}^n \right) &\leq \Delta_n \frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left(\sup_{u \in [0, \Delta_n]} |f(X_{t_{i-1}^n + u}) - f(X_{t_{i-1}^n})|^2 \,\middle|\, \mathcal{F}_{i-1}^n \right) \\
&= \Delta_n^2 \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}; \theta_0) \xrightarrow{\mathbb{P}_0} 0.
\end{aligned}$$

The desired conclusion (2.49) now follows immediately from Lemma 9 in Genon-Catalot and Jacod (1993).

Furthermore, Proposition 2.3.3 and Itô's formula applied to $U_0(f)$ imply that

$$\begin{aligned} U_0(f)(X_t) &= U_0(f)(X_0) + \int_0^t \mathcal{L}_0(U_0(f))(X_s)\,\mathrm{d}s + \int_0^t \partial_x U_0(f)(X_s) b(X_s;\theta_0) dB_s \\ &= U_0(f)(X_0) - \int_0^t f(X_s)\,\mathrm{d}s + \int_0^t \partial_x U_0(f)(X_s) b(X_s;\theta_0) dB_s, \end{aligned}$$

and, since $U_0(f) \in \mathscr{L}^2(\mu_0)$,

$$\frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} f(X_s)\,\mathrm{d}s = \frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} \partial_x U_0(f)(X_s) b(X_s;\theta_0) dB_s + o_{\mathbb{P}_0}(1). \tag{2.50}$$

The stochastic integral on the r.h.s. of (2.50) is easily shown to be a true martingale under $\mathbb{P}_0$ and by the pointwise ergodic theorem

$$\frac{1}{n\Delta_n} \int_0^{n\Delta_n} \left[\partial_x U_0(f)(X_s) b(X_s;\theta_0)\right]^2 \,\mathrm{d}s \xrightarrow{\mathbb{P}_0} \mu_0\left([\partial_x U_0(f) b(\,\cdot\,;\theta_0)]^2\right).$$

Collecting our observations,

$$\begin{aligned} \sqrt{n\Delta_n} V_n(f) &= \frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} f(X_s)\,\mathrm{d}s + o_{\mathbb{P}_0}(1) && (2.51) \\ &= \frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} \partial_x U_0(f)(X_s) b(X_s;\theta_0) dB_s + o_{\mathbb{P}_0}(1) && (2.52) \\ &\xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mu_0\left([\partial_x U_0(f) b(\,\cdot\,;\theta_0)]^2\right)\right), \end{aligned}$$

where convergence in law under $\mathbb{P}_0$ follows from the continuous-time martingale central limit theorem; see e.g. Theorem 6.31, Häusler and Luschgy (2015).

The alternative expression for the asymptotic variance $\mathcal{V}_0(f)$ in (2.16) follows from the result that

$$\frac{1}{\sqrt{T}} \int_0^T f(X_s)\,\mathrm{d}s \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mu_0(f U_0(f))\right), \tag{2.53}$$

as $T \to \infty$ for any $f \in \mathscr{L}_0^2(\mu_0)$ and, hence, in particular for $f \in \mathscr{H}_0^2$; see e.g. Theorem 2.2 in Genon-Catalot, Jeantheau, and Larédo (2000). The reader can consult Lemma VIII.3.68 in Jacod and Shiryaev (2003) for details about the proof. □

Proof of Theorem 2.4.2. To prove the eventual existence of a consistent sequence of G_n-estimators $(\hat{\theta}_n)$, we argue that the equivalent estimating function

$$H_n(\theta) = \frac{1}{n} \sum_{i=1}^n \left[f(X_{t_i^n}) - \mu_\theta(f)\right]$$

satisfies the regularity conditions of Theorem 1.58 in M. Sørensen (2012); by Lemma 2.3.1,

$$H_n(\theta) \xrightarrow{\mathbb{P}_0} H(\theta) = (\mu_0 - \mu_\theta)(f),$$

and since $\partial_\theta H_n(\theta) = -\partial_\theta \mu_\theta(f)$ for every $n \in \mathbb{N}$,

$$\sup_{\theta \in \Theta} |\partial_\theta H_n(\theta) + \partial_\theta \mu_\theta(f)| = 0,$$

and existence follows from continuity of the map $\theta \mapsto \partial_\theta H_n(\theta)$ and the additional assumption $\partial_\theta \mu_\theta(f) \neq 0$.

Moreover, if we denote by $\bar{B}_\varepsilon(\theta_0)$ the closed ball with radius $\varepsilon > 0$ centered at θ_0, the identifiability assumption $H(\theta) \neq 0$ for all $\theta \neq \theta_0$ together with continuity of $\theta \mapsto H(\theta)$ implies that

$$\mathbb{P}_0 \left(\inf_{\theta \in \mathcal{K} \setminus \bar{B}_\varepsilon(\theta_0)} |H(\theta)| > 0 \right) = 1$$

for arbitrary $\varepsilon > 0$ and any compact subset $\mathcal{K} \subset \Theta$ that contains θ_0. Hence, by Theorem 1.59 in M. Sørensen (2012), any sequence of G_n-estimators must either be consistent or converge to the boundary of Θ, $\mathbb{P}_0$-almost surely. In particular, we conclude that $\hat{\theta}_n$ will be unique in any compact subset $\mathcal{K} \subset \Theta$ that contains θ_0 with $\mathbb{P}_0$-probability approaching one as $n \to \infty$.

To establish asymptotic normality, a first order Taylor expansion of H_n implies that

$$0 = \sqrt{n\Delta_n} H_n(\theta_0) + \partial_\theta H_n(\theta_n^*) \sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0) \tag{2.54}$$

for some θ_n^* between $\hat{\theta}_n$ and θ_0 and, by Proposition 2.3.4,

$$\begin{aligned}
\sqrt{n\Delta_n} H_n(\theta_0) &= \sqrt{n\Delta_n} V_n(f^*) + \left(\frac{\Delta_n}{n}\right)^{1/2} [f^*(X_{n\Delta_n}) - f^*(X_0)] \\
&= \sqrt{n\Delta_n} V_n(f^*) + o_{\mathbb{P}_0}(1) \\
&\xrightarrow{\mathscr{D}_0} \mathcal{N}(0, \mathcal{V}_0(f)),
\end{aligned}$$

where

$$\mathcal{V}_0(f) = 2 \int_S f^*(x) U_0(f^*)(x) \mu_0(dx).$$

Since $\partial_\theta H_n(\theta_n^*) \xrightarrow{\mathbb{P}_0} -\partial_\theta \mu_0(f)$ as $n \to \infty$, the continuous mapping theorem applied to (2.54) yields the result. □

Proof of Lemma 2.4.4. For simplicity, we define

$$H_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \tag{2.55}$$

where $g = (g_1, g_2)^T$ is given by

$$g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) = f(X_{t_i^n}) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}), \tag{2.56}$$

$$g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) = f(X_{t_{i-1}^n}) \left[f(X_{t_i^n}) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}) \right], \tag{2.57}$$

corresponding to the entries of G_n.

Our first step will be to verify the coefficient expansion (2.21) of $\breve{a}_n(\theta)$ into powers of Δ_n. By applying Lemma 2.10.2,

$$\mathbb{E}_\theta \left(f(X_{\Delta_n}) \mid \mathcal{F}_0 \right) = f(X_0) + \Delta_n \mathcal{L}_\theta f(X_0) + \Delta_n^2 R(\Delta_n, X_0; \theta),$$

which implies that

$$\begin{aligned} \mathbb{E}_\theta \left[f(X_0) f(X_{\Delta_n}) \right] &= \mathbb{E}_\theta \left[f(X_0) \mathbb{E}_\theta (f(X_{\Delta_n}) \mid \mathcal{F}_0) \right] \\ &= \mu_\theta(f^2) + \Delta_n \mu_\theta(f \mathcal{L}_\theta f) + \Delta_n^2 R(\Delta_n; \theta), \end{aligned}$$

where $|R(\Delta_n; \theta)| \leq C(\theta)$ for a constant $C > 0$. This yields the Δ_n-expansion

$$\breve{a}_n(\theta)_1 = \frac{\mathbb{E}_\theta \left[f(X_0) f(X_{\Delta_n}) \right] - \left[\mu_\theta(f) \right]^2}{\mathbb{V}ar_\theta f(X_0)} = 1 + \Delta_n K_f(\theta) + \Delta_n^2 R(\Delta_n; \theta), \tag{2.58}$$

and, as a consequence,

$$\breve{a}_n(\theta)_0 = -\Delta_n K_f(\theta) \mu_\theta(f) + \Delta_n^2 R(\Delta_n; \theta). \tag{2.59}$$

In turn, this enables us to expand $\mathbb{E}_0 \left[g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \;\middle|\; \mathcal{F}_{i-1}^n \right]$ into powers of Δ_n since

$$\mathbb{E}_0 \left(f(X_{t_i^n}) \;\middle|\; \mathcal{F}_{i-1}^n \right) = f(X_{t_{i-1}^n}) + \Delta_n \mathcal{L}_0 f(X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta_0),$$

which together with (2.58) and (2.59) implies that

$$\begin{aligned} & \mathbb{E}_0 \left[g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \;\middle|\; \mathcal{F}_{i-1}^n \right] \\ = {} & \mathbb{E}_0 \left(f(X_{t_i^n}) \;\middle|\; \mathcal{F}_{i-1}^n \right) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}) \\ = {} & \Delta_n \left(\mathcal{L}_0 f(X_{t_{i-1}^n}) + K_f(\theta) \left[\mu_\theta(f) - f(X_{t_{i-1}^n}) \right] \right) + \Delta_n^2 R_0(\Delta_n, X_{t_{i-1}^n}; \theta). \end{aligned} \tag{2.60}$$

Hence, by Lemma 2.3.1,

$$\begin{aligned} & \frac{1}{n\Delta_n} \sum_{i=1}^n \mathbb{E}_0 \left[g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \;\middle|\; \mathcal{F}_{i-1}^n \right] \\ = {} & \frac{1}{n} \sum_{i=1}^n \mathcal{L}_0 f(X_{t_{i-1}^n}) + K_f(\theta) \cdot \frac{1}{n} \sum_{i=1}^n \left[\mu_\theta(f) - f(X_{t_{i-1}^n}) \right] + \frac{\Delta_n}{n} \sum_{i=1}^n R_0(\Delta_n, X_{t_{i-1}^n}; \theta) \\ \xrightarrow{\mathbb{P}_0} {} & K_f(\theta)(\mu_\theta - \mu_0)(f), \end{aligned}$$

where, in particular, the contribution from the first term vanishes since $\mu_0(\mathcal{L}_0 f) = 0$; see e.g. Hansen and Scheinkman (1995).

To apply Lemma 9 in Genon-Catalot and Jacod (1993), it remains to show that

$$\frac{1}{n^2\Delta_n^2}\sum_{i=1}^n \mathbb{E}_0\left[g_1^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,\middle|\, \mathcal{F}_{i-1}^n\right] = o_{\mathbb{P}_0}(1). \tag{2.61}$$

From the coefficient expansions (2.58) and (2.59), it follows that

$$\breve{\pi}_{i-1}(\theta) = \breve{a}_n(\theta)_0 + \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}) = f(X_{t_{i-1}^n}) + \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta),$$

which, in turn, yields the decomposition

$$\begin{aligned} g_1^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) = \left[f(X_{t_i^n}) - f(X_{t_{i-1}^n})\right]^2 + \\ \left[f(X_{t_i^n}) - f(X_{t_{i-1}^n})\right]\Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta). \end{aligned} \tag{2.62}$$

For the first term, Lemma 2.10.1 implies that

$$\frac{1}{n^2\Delta_n^2}\sum_{i=1}^n \mathbb{E}_0\left[|f(X_{t_i^n}) - f(X_{t_{i-1}^n})|^2\,\middle|\, \mathcal{F}_{i-1}^n\right] = \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta_0) \xrightarrow{\mathbb{P}_0} 0,$$

where we apply that $n\Delta_n \to \infty$. Similarly,

$$\begin{aligned} & \frac{1}{n^2\Delta_n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta)\mathbb{E}_0\left[|f(X_{t_i^n}) - f(X_{t_{i-1}^n})|\,\middle|\, \mathcal{F}_{i-1}^n\right] \\ = \ & \frac{\Delta_n^{1/2}}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R_0(\Delta_n, X_{t_{i-1}^n};\theta) \\ = \ & o_{\mathbb{P}_0}(1), \end{aligned}$$

and, finally,

$$\frac{1}{n^2}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta) \xrightarrow{\mathbb{P}_0} 0,$$

which together implies (2.61). Thus,

$$\frac{1}{n\Delta_n}\sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \xrightarrow{\mathbb{P}_0} K_f(\theta)(\mu_\theta - \mu_0)(f).$$

Similarly for $g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)$, it follows easily from (2.60) that

$$\begin{aligned} & \mathbb{E}_0\left[g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,\middle|\, \mathcal{F}_{i-1}^n\right] \\ = \ & \Delta_n\left(f(X_{t_{i-1}^n})\mathcal{L}_0 f(X_{t_{i-1}^n}) - K_f(\theta) f(X_{t_{i-1}^n})\left[f(X_{t_{i-1}^n}) - \mu_\theta(f)\right]\right) \\ + \ & \Delta_n^2 R_0(\Delta_n, X_{t_{i-1}^n};\theta), \end{aligned}$$

and, hence,

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n}\mathbb{E}_0\left[g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,\middle|\,\mathcal{F}_{i-1}^n\right] \xrightarrow{\mathbb{P}_0} \mu_0(f\mathcal{L}_0 f) - K_f(\theta)\left[\mu_0(f^2) - \mu_0(f)\mu_\theta(f)\right].$$

Moreover, since $g_2^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) = f^2(X_{t_{i-1}^n})g_1^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)$, we easily see that

$$\frac{1}{n^2\Delta_n^2}\sum_{i=1}^{n}\mathbb{E}_0\left[g_2^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,\middle|\,\mathcal{F}_{i-1}^n\right] = o_{\mathbb{P}_0}(1).$$

To establish the limit of $\partial_{\theta^T}H_n(\theta)$, we write

$$H_n(\theta) = \frac{1}{n\Delta_n}\sum_{i=1}^{n} Z_{i-1}\left[f(X_{t_i^n}) - Z_{i-1}^T\breve{a}_n(\theta)\right],$$

which shows that

$$\partial_{\theta^T}H_n(\theta) = -\frac{1}{n\Delta_n}\sum_{i=1}^{n} Z_{i-1}Z_{i-1}^T\partial_{\theta^T}\breve{a}_n(\theta) = Z_n(f)A_n(\theta),$$

where $Z_n(f) := \frac{1}{n}\sum_{i=1}^n Z_{i-1}Z_{i-1}^T$ and $A_n(\theta) := -\Delta_n^{-1}\partial_{\theta^T}\breve{a}_n(\theta)$. By Lemma 2.3.1,

$$Z_n(f) \xrightarrow{\mathbb{P}_0} Z(f) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix}$$

and applying the power expansion (2.21),

$$A_n(\theta) = \partial_{\theta^T}\begin{pmatrix} K_f(\theta)\mu_\theta(f) \\ -K_f(\theta) \end{pmatrix} + \Delta_n\partial_{\theta^T}R(\Delta_n;\theta) \to \partial_{\theta^T}\begin{pmatrix} K_f(\theta)\mu_\theta(f) \\ -K_f(\theta) \end{pmatrix} =: A(\theta),$$

which holds for all $\theta\in\mathcal{M}$ under the regularity assumption (2.23). Collecting our observations,

$$\partial_{\theta^T}H_n(\theta) \xrightarrow{\mathbb{P}_0} Z(f)A(\theta) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix}\begin{pmatrix} \partial_{\theta_1}\left[K_f(\theta)\mu_\theta(f)\right] & \partial_{\theta_2}\left[K_f(\theta)\mu_\theta(f)\right] \\ -\partial_{\theta_1}K_f(\theta) & -\partial_{\theta_2}K_f(\theta) \end{pmatrix}.$$

To argue that the convergence is uniform over $\mathcal{M}$, note that

$$\begin{aligned} \|\partial_{\theta^T}H_n(\theta) - Z(f)A(\theta)\| &= \|Z_n(f)A_n(\theta) - Z(f)A(\theta)\| \\ &= \|Z_n(f)A_n(\theta) - Z_n(f)A(\theta) + Z_n(f)A(\theta) - Z(f)A(\theta)\| \\ &\leq \|Z_n(f)[A_n(\theta) - A(\theta)]\| + \|[Z_n(f) - Z(f)]A(\theta)\| \end{aligned}$$

and, in particular,

$$\sup_{\theta\in\mathcal{M}}\|\partial_{\theta^T}H_n(\theta) - Z(f)A(\theta)\| \leq$$
$$\|Z_n(f)\|\sup_{\theta\in\mathcal{M}}\|A_n(\theta) - A(\theta)\| + \|Z_n(f) - Z(f)\|\sup_{\theta\in\mathcal{M}}\|A(\theta)\|.$$

By continuity of norms, $\|Z_n(f)\| \xrightarrow{\mathbb{P}_0} \|Z(f)\|$, $\|Z_n(f) - Z(f)\| = o_{\mathbb{P}_0}(1)$ and (2.24) follows by observing that

$$\sup_{\theta\in\mathcal{M}} \|A_n(\theta) - A(\theta)\| = \Delta_n \sup_{\theta\in\mathcal{M}} \|\partial_{\theta^T} R(\Delta_n;\theta)\| \leq_{C(\mathcal{M})} \Delta_n \to 0$$

and applying continuity of $\theta \mapsto A(\theta)$. □

Proof of Theorem 2.4.5. We continue with the notation (2.55)-(2.57) introduced above. Existence of a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ follows from Lemma 2.4.4 and Theorem 1.58 in M. Sørensen (2012). Moreover, by the same reasoning as in the proof of Theorem 2.4.2, the identifiability assumption $\gamma(\theta_0;\theta) \neq 0$ for $\theta \neq \theta_0$ and continuity of $\theta \mapsto \gamma(\theta_0;\theta)$ imply that the solution $\hat{\theta}_n$ will be unique in any compact subset $\mathcal{K} \subset \Theta$ that contains θ_0 with $\mathbb{P}_0$-probability approaching one as $n \to \infty$.

To establish asymptotic normality for $n\Delta_n^3 \to 0$, we again rely on a first order Taylor expansion

$$0 = \sqrt{n\Delta_n} H_n(\theta_0) + \partial_{\theta^T} H_n(\theta_n^*) \sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0),$$

where θ_n^* lies between $\hat{\theta}_n$ and θ_0, $\partial_{\theta^T} H_n(\theta_n^*) \xrightarrow{\mathbb{P}_0} W(\theta_0)$ and the main difficulty is to show that

$$\sqrt{n\Delta_n} H_n(\theta_0) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2(0, \mathcal{V}_0(f)).$$

Reusing the coefficient expansions (2.58) and (2.59), we find that

$$g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) = f(X_{t_i^n}) - f(X_{t_{i-1}^n}) + \Delta_n f_1^*(X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta_0)$$

where f_1^* is defined in Condition 2.4.3. Hence,

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) =$$
$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \left[f(X_{t_i^n}) - f(X_{t_{i-1}^n})\right] + \sqrt{n\Delta_n} \cdot V_n(f_1^*) + (n\Delta_n^3)^{1/2} \cdot \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta_0)$$

and recognizing the first term as a telescoping sum, Proposition 2.3.4 implies that

$$\begin{aligned} \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \quad &= \quad \sqrt{n\Delta_n} \cdot V_n(f_1^*) + o_{\mathbb{P}_0}(1) \\ &\xrightarrow{\mathscr{D}_0} \quad \mathcal{N}\left(0, \mu_0\left([\partial_x U_0(f_1^*) b(\,\cdot\,;\theta_0)]^2\right)\right). \end{aligned}$$

The second entry of $\sqrt{n\Delta_n} \cdot H_n(\theta_0)$ requires a bit more work; by Itô's formula,

$$f(X_{t_i^n}) - f(X_{t_{i-1}^n}) = \Delta_n \mathcal{L}_0 f(X_{t_{i-1}^n}) + A_i(\theta_0) + M_i(\theta_0),$$

where

$$\begin{aligned} A_i(\theta) &= \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\mathcal{L}_\theta f(X_s) - \mathcal{L}_\theta f(X_{t_{i-1}^n})\right] \mathrm{d}s, \\ M_i(\theta) &= \int_{(i-1)\Delta_n}^{i\Delta_n} \partial_x f(X_s) b(X_s;\theta) dB_s, \end{aligned}$$

and, hence, by applying the expansions (2.58) and (2.59) as above,

$$g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) =$$
$$f(X_{t_{i-1}^n})A_i(\theta_0) + \Delta_n f_2^*(X_{t_{i-1}^n}) + f(X_{t_{i-1}^n})M_i(\theta_0) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta_0).$$

A straightforward extension of the proof of (2.49) implies that

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n f(X_{t_{i-1}^n})A_i(\theta_0) = o_{\mathbb{P}_0}(1)$$

since $n\Delta_n^3 \to 0$ and, as a consequence,

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) = \sqrt{n\Delta_n} V_n(f_2^*) + \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n f(X_{t_{i-1}^n}) M_i^n(\theta_0) + o_{\mathbb{P}_0}(1).$$

To gather the non-negligible terms, we argue as in (2.51)-(2.52) that

$$\begin{aligned} &\sqrt{n\Delta_n} V_n(f_2^*) \\ = &\frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} f_2^*(X_s)\,\mathrm{d}s + o_{\mathbb{P}_0}(1) \\ = &\frac{1}{\sqrt{n\Delta_n}} \int_0^{n\Delta_n} \partial_x U_0(f_2^*)(X_s) b(X_s;\theta_0) dB_s + o_{\mathbb{P}_0}(1) \\ = &\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \partial_x U_0(f_2^*)(X_s) b(X_s;\theta_0) dB_s + o_{\mathbb{P}_0}(1), \end{aligned}$$

which, in turn, yields the stochastic integral representation

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) =$$
$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\partial_x U_0(f_2^*)(X_s) + f(X_{t_{i-1}^n}) \partial_x f(X_s)\right] b(X_s;\theta_0) dB_s + o_{\mathbb{P}_0}(1).$$

At this point, we can apply the CLT for martingale difference arrays; see e.g. Häusler and Luschgy (2015). To shorten notation in the following, we let

$$Z_i := \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\partial_x U_0(f_2^*)(X_s) + f(X_{t_{i-1}^n}) \partial_x f(X_s)\right] b(X_s;\theta_0) dB_s,$$

and introduce

$$h(x) = \left[\partial_x U_0(f_2^*)(x) + f(x)\partial_x f(x)\right]^2 b^2(x;\theta_0).$$

Firstly, by the conditional Itô isometry, Tonelli's theorem and Lemma 2.10.2,

$$\begin{aligned}
& \frac{1}{n\Delta_n}\sum_{i=1}^n \mathbb{E}_0\left((Z_i)^2 \;\middle|\; \mathcal{F}_{i-1}^n\right) \\
= \;& \frac{1}{n\Delta_n}\sum_{i=1}^n \mathbb{E}_0\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \left[\partial_x U_0(f_2^*)(X_s) + f(X_{t_{i-1}^n})\partial_x f(X_s)\right]^2 b^2(X_s;\theta_0)\,\mathrm{d}s \;\middle|\; \mathcal{F}_{i-1}^n\right) \\
= \;& \frac{1}{n\Delta_n}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}_0\left(\left[\partial_x U_0(f_2^*)(X_s) + f(X_{t_{i-1}^n})\partial_x f(X_s)\right]^2 b^2(X_s;\theta_0) \;\middle|\; \mathcal{F}_{i-1}^n\right)\mathrm{d}s \\
= \;& \frac{1}{n\Delta_n}\sum_{i=1}^n \int_0^{\Delta_n} \left[h(X_{t_{i-1}^n}) + u\cdot R(u, X_{t_{i-1}^n};\theta_0)\right]\mathrm{d}u \\
= \;& \frac{1}{n}\sum_{i=1}^n h(X_{t_{i-1}^n}) + o_{\mathbb{P}_0}(1) \\
\xrightarrow{\mathbb{P}_0} \;& \mu_0\left(\left[\partial_x U_0(f_2^*) + f\partial_x f\right]^2 b^2(\,\cdot\,;\theta_0)\right).
\end{aligned}$$

Moreover, for any $g \in \mathcal{C}_p^2(S)$ and $k \geq 2$, the Burkholder-Davis-Gundy inequality, Jensen's inequality, Tonelli's theorem and Lemma 2.10.2, respectively, imply that

$$\begin{aligned}
& \mathbb{E}_0\left(\left|\int_{(i-1)\Delta_n}^{i\Delta_n} g(X_s)dB_s\right|^k \;\middle|\; \mathcal{F}_{i-1}^n\right) \\
\leq \;& \mathbb{E}_0\left(\left(\int_{(i-1)\Delta_n}^{i\Delta_n} g^2(X_s)\,\mathrm{d}s\right)^{k/2} \;\middle|\; \mathcal{F}_{i-1}^n\right) \\
\leq \;& \Delta_n^{k/2-1}\cdot\mathbb{E}_0\left(\int_{(i-1)\Delta_n}^{i\Delta_n} |g(X_s)|^k\,\mathrm{d}s \;\middle|\; \mathcal{F}_{i-1}^n\right) \\
= \;& \Delta_n^{k/2-1}\cdot\int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}_0\left(|g(X_s)|^k \;\middle|\; \mathcal{F}_{i-1}^n\right)\mathrm{d}s \\
= \;& \Delta_n^{k/2-1}\cdot\int_0^{\Delta_n} \left(|g(X_{t_{i-1}^n})|^k + u\cdot R(u, X_{t_{i-1}^n};\theta_0)\right)\mathrm{d}u \\
\leq_C \;& \Delta_n^{k/2}|g(X_{t_{i-1}^n})|^k + \Delta_n^{k/2+1}F(X_{t_{i-1}^n};\theta_0),
\end{aligned}$$

so based on the inequality

$$\begin{aligned}
& |Z_i|^3 \leq_C \\
& \left|\int_{(i-1)\Delta_n}^{i\Delta_n} \partial_x U_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s\right|^3 + |f(X_{t_{i-1}^n})|^3\left|\int_{(i-1)\Delta_n}^{i\Delta_n} \partial_x f(X_s)b(X_s;\theta_0)dB_s\right|^3,
\end{aligned}$$

we conclude that

$$\begin{aligned}
&\frac{1}{(n\Delta_n)^{3/2}}\sum_{i=1}^{n}\mathbb{E}_0\left(|Z_i|^3 \mid \mathcal{F}_{i-1}^n\right)\\
\leq_C\quad &\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\left[|\partial_x U_0(f_2^*)(X_{t_{i-1}^n})|^3+|f(X_{t_{i-1}^n})|^3|\partial_x f(X_{t_{i-1}^n})|^3\right]|b(X_{t_{i-1}^n};\theta_0)|^3+o_{\mathbb{P}_0}(1)\\
\xrightarrow{\mathbb{P}_0}\quad & 0.
\end{aligned}$$

By the martingale CLT for triangular arrays,

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}g_2(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta_0)\xrightarrow{\mathscr{D}_0}\mathcal{N}\left(0,\mu_0\left(\left[\partial_x U_0(f_2^*)+f\partial_x f\right]^2 b^2(\,\cdot\,;\theta_0)\right)\right). \tag{2.63}$$

Finally, we apply the Cramér-Wold device to establish joint convergence in law; using the martingale decomposition

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}g_1(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta_0)=\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}\int_{(i-1)\Delta_n}^{i\Delta_n}\partial_x U_0(f_1^*)(X_s)b(X_s;\theta_0)dB_s+o_{\mathbb{P}_0}(1),$$

a minor modification of the proof of (2.63) shows that

$$\begin{aligned}
&c_1\cdot\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}g_1(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta_0)+c_2\cdot\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}g_2(\Delta_n,X_{t_i^n},X_{t_{i-1}^n};\theta_0)\\
=\quad&\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\left(\partial_x U_0(c_1f_1^*+c_2f_2^*)(X_s)+c_2f(X_{t_{i-1}^n})\partial_x f(X_s)\right)b(X_s;\theta_0)dB_s+o_{\mathbb{P}_0}(1)\\
\xrightarrow{\mathscr{D}_0}\quad&\mathcal{N}\left(0,\mu_0\left(\left[\partial_x U_0(c_1f_1^*+c_2f_2^*)+c_2f\partial_x f\right]^2 b^2(\,\cdot\,;\theta_0)\right)\right)
\end{aligned}$$

for arbitrary constants $c_1, c_2 \in \mathbb{R}$. In particular, we apply linearity of the potential operator $f \mapsto U_0(f)$ to write

$$c_1\partial_x U_0(f_1^*)+c_2\partial_x U_0(f_2^*)=\partial_x U_0(c_1f_1^*+c_2f_2^*).$$

□

Proof of Proposition 2.5.2. By the general inequality (2.48),

$$\|U_\theta(g)\|_2\leq\lambda^{-1}\|g\|_2$$

where $\lambda>0$ denotes the spectral gap of (X_t) under $\mathbb{P}_\theta$ and $g\in\mathscr{H}_\theta^2$ is arbitrary. Therefore, by the Cauchy-Schwarz inequality,

$$|\mu_\theta\left(gU_\theta(g)\right)|\leq\|g\|_2\|U_\theta(g)\|_2\leq\frac{\|g\|_2^2}{\lambda}$$

and, as a special case,

$$\mathrm{AVAR}(\hat{\theta}_n)=\frac{2\mu_0\left(f^*U_0(f^*)\right)}{[\partial_\theta\mu_0(f)]^2}\leq\left(\frac{2}{\lambda_0}\right)\frac{\mathbb{V}ar_0 f(X_0)}{[\partial_\theta\mu_0(f)]^2}.$$

□

2.10 Appendix B: Moment expansions

The proofs in Section 2.9 rely on conditional moment expansions for diffusion models and the following results are essentially taken from Gloter (2000) and Florens-Zmirou (1989), respectively. Rough proofs are provided for completeness. In the sequel, $\theta \in \Theta$ is arbitrary and we assume for convenience that $0 < \Delta < 1$.

Lemma 2.10.1. *Let $f \in \mathcal{C}_p^1(S)$. For any $k \geq 1$,*

$$\mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} |f(X_{t+s}) - f(X_t)|^k \,\middle|\, \mathcal{F}_t \right) \leq_{C_k} \Delta^{k/2} \left(1 + |X_t|\right)^{C_k}.$$

Proof. We start by showing the result for $f(x) = x$ and since

$$\mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} |X_{t+s} - X_t| \,\middle|\, \mathcal{F}_t \right) \leq \mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} |X_{t+s} - X_t|^2 \,\middle|\, \mathcal{F}_t \right)^{1/2}$$

by Hölder's inequality, it suffices to consider $k \geq 2$.

Firstly, since

$$|X_{t+s} - X_t| = \left| \int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u + \int_t^{t+s} b(X_u;\theta) dB_u \right|,$$

we easily derive that

$$\mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} |X_{t+s} - X_t|^k \,\middle|\, \mathcal{F}_t \right) \leq_C$$
$$\mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} \left| \int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u \right|^k \,\middle|\, \mathcal{F}_t \right) + \mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} \left| \int_t^{t+s} b(X_u;\theta) dB_u \right|^k \,\middle|\, \mathcal{F}_t \right).$$

To bound the drift component, it follows by Jensen's inequality that

$$\begin{aligned} \mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} \left| \int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u \right|^k \,\middle|\, \mathcal{F}_t \right) &= \mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} s^k \left| \frac{1}{s} \int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u \right|^k \,\middle|\, \mathcal{F}_t \right) \\ &\leq \mathbb{E}_\theta \left(\sup_{s \in [0,\Delta]} s^{k-1} \int_t^{t+s} |a(X_u;\theta)|^k \,\mathrm{d}u \,\middle|\, \mathcal{F}_t \right) \end{aligned}$$

and, using the linear growth assumption of $a(x;\theta)$ in Condition 2.2.2,

$$|a(X_u;\theta)|^k \leq_C (1 + |X_u|)^k \leq_C 1 + |X_t|^k + |X_u - X_t|^k,$$

which yields the upper bound

$$
\begin{aligned}
& \mathbb{E}_\theta\left(\sup_{s\in[0,\Delta]}\left|\int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u\right|^k \,\middle|\, \mathcal{F}_t\right) \\
\leq_C \quad & \mathbb{E}_\theta\left(\sup_{s\in[0,\Delta]} s^{k-1}\int_t^{t+s}\left(1+|X_t|^k+|X_u-X_t|^k\right)\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
= \quad & \mathbb{E}_\theta\left(\Delta^{k-1}\int_t^{t+\Delta}\left(1+|X_t|^k+|X_u-X_t|^k\right)\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
= \quad & \Delta^k(1+|X_t|^k)+\Delta^{k-1}\int_t^{t+\Delta}\mathbb{E}_\theta\left(|X_u-X_t|^k \,\middle|\, \mathcal{F}_t\right)\mathrm{d}u \\
\leq \quad & \Delta^{k/2}(1+|X_t|^k)+\int_0^\Delta \mathbb{E}_\theta\left(\sup_{s\in[0,u]}|X_{t+s}-X_t|^k \,\middle|\, \mathcal{F}_t\right)\mathrm{d}u, \qquad (2.64)
\end{aligned}
$$

where we apply Tonelli's theorem to interchange the order of integration.

By similar reasoning, we bound the local martingale component using Jensen's inequality and a conditional version of the Burkholder-Davis-Gundy inequalities; see e.g. Jacod and Protter (2012). For $k \geq 2$,

$$
\begin{aligned}
& \mathbb{E}_\theta\left(\sup_{s\in[0,\Delta]}\left|\int_t^{t+s} b(X_u;\theta)dB_u\right|^k \,\middle|\, \mathcal{F}_t\right) \\
\leq_C \quad & \mathbb{E}_\theta\left(\left|\int_t^{t+\Delta} b^2(X_u;\theta)\,\mathrm{d}u\right|^{k/2} \,\middle|\, \mathcal{F}_t\right) \\
\leq \quad & \Delta^{k/2-1}\cdot\mathbb{E}_\theta\left(\int_t^{t+\Delta}|b(X_u;\theta)|^k\,\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
\leq_C \quad & \Delta^{k/2-1}\cdot\mathbb{E}_\theta\left(\int_t^{t+\Delta}\left(1+|X_t|^k+|X_u-X_t|^k\right)\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
\leq \quad & \Delta^{k/2}(1+|X_t|^k)+\int_0^\Delta \mathbb{E}_\theta\left(\sup_{s\in[0,u]}|X_{t+s}-X_t|^k \,\middle|\, \mathcal{F}_t\right)\mathrm{d}u. \qquad (2.65)
\end{aligned}
$$

Since the upper bounds (2.64) and (2.65) coincide, we conclude that

$$
\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta]}|X_{t+s}-X_t|^k \,\middle|\, \mathcal{F}_t\right) \leq_C \Delta^{k/2}(1+|X_t|^k)+\int_0^\Delta \mathbb{E}_\theta\left(\sup_{s\in[0,u]}|X_{t+s}-X_t|^k \,\middle|\, \mathcal{F}_t\right)\mathrm{d}u
$$

and, at this point, the desired result

$$
\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta]}|X_{t+s}-X_t|^k \,\middle|\, \mathcal{F}_t\right) \leq_C \Delta^{k/2}(1+|X_t|)^C, \qquad (2.66)
$$

holds by a Grönwall-type inequality; see e.g. Theorem 1.3 in Bainov and Simeonov (1992).

For general f, the first order Taylor expansion

$$f(X_{t+s}) = f(X_t) + \partial_x f(X^*)(X_{t+s} - X_t),$$

is satisfied for some X^* between X_t and X_{t+s}. Hence,

$$\mathbb{E}_\theta \left(\sup_{s\in[0,\Delta]} |f(X_{t+s}) - f(X_t)|^k \,\middle|\, \mathcal{F}_t \right) = \mathbb{E}_\theta \left(\sup_{s\in[0,\Delta]} |\partial_x f(X^*)(X_{t+s} - X_t)|^k \,\middle|\, \mathcal{F}_t \right)$$

and since $\partial_x f$ is of polynomial growth,

$$\sup_{s\in[0,\Delta]} |\partial_x f(X^*)|^k \leq_{C_k} 1 + |X_t|^{C_k} + \sup_{s\in[0,\Delta]} |X_{t+s} - X_t|^{C_k}$$

and a double application of (2.66) yields the result. □

Lemma 2.10.2. *Suppose that* $a(x;\theta) \in \mathcal{C}_p^{2k,0}(S\times\Theta)$, $b(x;\theta) \in \mathcal{C}_p^{2k,0}(S\times\Theta)$ *and* $f \in \mathcal{C}_p^{2(k+1)}(S)$ *for some* $k \geq 0$. *Then,*

$$\mathbb{E}_\theta \left(f(X_{t+\Delta}) \mid \mathcal{F}_t \right) = \sum_{i=0}^{k} \frac{\Delta^i}{i!} \mathcal{L}_\theta^i f(X_t) + \Delta^{k+1} R(\Delta, X_t; \theta).$$

Proof. We only consider $k = 0$, the general case may be shown by induction; see Lemma 1.10, M. Sørensen (2012). By Itô's formula,

$$f(X_{t+\Delta}) = f(X_t) + \int_t^{t+\Delta} \mathcal{L}_\theta f(X_s)\,\mathrm{d}s + \int_t^{t+\Delta} \partial_x f(X_s) b(X_s;\theta) dB_s,$$

and since $\partial_x f$ and $b(\,\cdot\,;\theta)$ are of polynomial, respectively linear, growth in x, the stochastic integral is a true $(\mathcal{F}_t)$-martingale w.r.t. $\mathbb{P}_\theta$ and

$$\mathbb{E}_\theta \left(f(X_{t+\Delta}) \mid \mathcal{F}_t \right) = f(X_t) + \int_0^{\Delta} \mathbb{E}_\theta \left(\mathcal{L}_\theta f(X_{t+u}) \mid \mathcal{F}_t \right) \mathrm{d}u.$$

Moreover, since $\mathcal{L}_\theta f$ is of polynomial growth in x,

$$|\mathcal{L}_\theta f(X_{t+u})| \leq_C 1 + |X_t|^C + |X_{t+u} - X_t|^C$$

and, hence,

$$\Delta^{-1} \int_0^{\Delta} \mathbb{E}_\theta \left(\mathcal{L}_\theta f(X_{t+u}) \mid \mathcal{F}_t \right) \mathrm{d}u = R(\Delta, X_t; \theta),$$

by a simple application of Lemma 2.10.1. □

3

Inference for Integrated Diffusions Observed at High Frequency

Emil S. Jørgensen and Michael Sørensen

University of Copenhagen

Abstract. Prediction-based estimating functions provide a general framework for parametric inference in discretized diffusion-type models. In this paper, we suppose we observe a discretization $\{I_{t_i^n}\}_{i=0}^n$ of an integral process $I_t = \int_0^t X_s \, \mathrm{d}s$, where (X_t) is a time-homogeneous diffusion with an unknown parameter $\theta \in \Theta \subset \mathbb{R}^d$ that we wish to estimate. The observation times $\{t_i^n\}$ are assumed to be deterministic and equidistant, i.e. $t_i^n = i\Delta_n$ for some $\Delta_n > 0$, and we consider the high-frequency asymptotic scenario where $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Subject to mild regularity conditions on (X_t), we prove existence of a consistent and asymptotically normal estimator $\hat{\theta}_n$ for a tractable class of prediction-based estimating functions. The proofs are based on power expansions for diffusion and integrated diffusion models and asymptotic normality is obtained under the additional rate assumption $n\Delta_n^2 \to 0$. Our results are of particular interest in finance, where realized volatility or variations thereof are often used to construct a trajectory of the latent integrated volatility process.

Keywords: *Euler-Itô expansion, high-frequency data, integrated diffusion, potential operator, prediction-based estimating functions, ρ-mixing.*

3.1 Introduction

Diffusion processes are widely used in many scientific areas, particularly in finance. While the processes are characterized in terms of continuous-time dynamics, available time series are always observed at discrete points in time. To bridge the gap between theory and applications, statistical methods for discretely observed (discretized) continuous-time stochastic processes is an active area of research today, and the recent availability of high-frequency data has spiked considerable interest into the construction of estimators and test statistics with nice asymptotic properties as the time between consecutive observations goes to zero.

This paper deals with parametric inference for integrated diffusion models $(I_t)_{t\geq 0}$ of the general form

$$\begin{aligned} dI_t &= X_t dt & (3.1)\\ dX_t &= a(X_t;\theta)dt + b(X_t;\theta)dB_t, & (3.2)\end{aligned}$$

where (X_t) takes values in an open interval $(l,r) \subset \mathbb{R}$ and the parameter of interest $\theta \in \Theta \subset \mathbb{R}^d$ for some $d \geq 1$. We suppose we observe a single discretization $\{I_{t_i^n}\}_{i=0}^n$ of the integrated process at deterministic, equidistant points in time and write $t_i^n = i\Delta_n$ for the appropriate $\Delta_n > 0$. To encompass consistent estimation of both drift and diffusion parameters, we consider the ergodic high-frequency sampling scenario

$$n \to \infty, \qquad \Delta_n \to 0, \qquad n \cdot \Delta_n \to \infty, \tag{3.3}$$

and suppose that the latent diffusion process (X_t) is strictly stationary under the probability measure $\mathbb{P}_\theta$ for an invariant distribution $X_0 \sim \mu_\theta$. A more appropriate, *equivalent* observation scheme is obtained for the transformed variables

$$Y_i = \Delta_n^{-1}\left(I_{t_i^n} - I_{t_{i-1}^n}\right) = \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} X_s \,\mathrm{d}s \tag{3.4}$$

where $i = 1, \ldots, n$. Note that for every fixed value of Δ_n, the sequence $\{Y_i\}_{i=1}^\infty$ inherits stationary under $\mathbb{P}_\theta$ from that of (X_t).

For the construction of estimators $\hat{\theta}_n$, we apply prediction-based estimating functions. This class of estimating functions was proposed by M. Sørensen (2000, 2011) as a versatile framework for parametric inference in non-Markovian diffusion-type models, and integrated diffusions were considered by Ditlevsen and M. Sørensen (2004) as a special case. As their main contribution, Ditlevsen and M. Sørensen (2004) illustrate a simple way to construct explicit Godambe optimal prediction-based estimating functions for (X_t) belonging to a tractable class of models that includes the Ornstein-Uhlenbeck process and the square-root process of Cox, Ingersoll, and Ross (1985). Low-frequency asymptotic results follow easily from general results in M. Sørensen (2000). The main contribution of this paper is a formal derivation of feasible high-frequency limit theorems for a large class of prediction-based estimating functions. Our proofs rely on the asymptotic results for diffusion models derived in Chapter 2 and we show that, under suitable regularity conditions, consistency and asymptotic normality is attained within the ergodic scenario (3.3).

Parametric estimation for discretely observed diffusion models (X_t) of the form (3.2) is the topic of many papers. Since the preferred method of maximum likelihood is infeasible for most models applied in practice, a wide range of alternative methods have been proposed and applied successfully. The Markov property of (X_t) enables most types of quasi-likelihood, including contrast functions (Yoshida (1992), Hansen and Scheinkman (1995), Kessler (1997)), estimating functions (Bibby and M. Sørensen (1995), Kessler (2000)), likelihood expansions (Dacunha-Castelle and Florens-Zmirou (1986), Aït-Sahalia (2002)), Markov-chain Monte Carlo (Elerian, Chib, and Shephard (2001), Roberts and Stramer (2001)) and simulated likelihood (Beskos, Papaspiliopoulos, Roberts, and Fearnhead (2006)) to name a few.

Although to a lesser extent, parametric inference for integrated diffusions has also been the topic of many papers in econometrics and statistics, the former in the guise of continuous-time stochastic volatility models. If we for illustrative purposes consider the simple stochastic volatility model

$$dS_t = \sqrt{v_t} dW_t, \tag{3.5}$$

where (W_t) denotes a standard Brownian motion, the availability of high-frequency observations of (S_t) enables us to filter out a trajectory of the latent integrated volatility

$$\int_0^t v_s \, \mathrm{d}s \tag{3.6}$$

and view it as an observable process.[1] This property has lead to the construction of estimators for integrated processes in the case where $v_t = v_t(\theta)$ for a parameter $\theta \in \Theta \subset \mathbb{R}^d$ that we wish to estimate, e.g. if the volatility dynamics are described by a time-homogeneous diffusion process

$$dv_t = a(v_t;\theta)dt + b(v_t;\theta)dB_t.$$

Examples of the latter include the GARCH(1,1) diffusion model in Nelson (1990), the square-root (CIR) process in Heston (1993) and the 3/2 diffusion in Drimus (2012). Estimation based on realized power variations that approximate the integrated volatility (3.6) have been studied by e.g. Bollerslev and Zhou (2002), Barndorff-Nielsen and Shephard (2002a) and Todorov (2009), the latter in a general GMM framework for a large class of stochastic volatility models with jumps. On a related note, the recent paper by J. Li and Xiu (2016) appears to be the first to develop high-frequency (infill) asymptotics for GMM estimators of parameters in the diffusion coefficient of the volatility process by preliminary filtering of the spot volatility instead. On the statistical side, the paper by Ditlevsen and M. Sørensen (2004) was summarized above and Baltazar-Larios and M. Sørensen (2010) propose a simulated EM-algorithm to obtain maximum likelihood estimators for integrated diffusions contaminated by (microstructure) noise. A third, and for this paper highly influential, approach based on expansion results for small values of Δ_n was proposed by Gloter (2000, 2006). His construction of contrast estimators utilizes the basic idea that, as $\Delta_n \to 0$,

$$Y_i = \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} X_s \, \mathrm{d}s \approx X_{t_{i-1}^n}$$

[1] Nonparametric filtering of integrated volatility from high-frequency financial time series is an emblematic problem in financial econometrics. The reader can consult e.g. Aït-Sahalia and Jacod (2014) for a recent monograph with an extensive list of references.

which allows for the derivation of high-frequency limit theorems for integrated diffusions.

Finally, nonparametric estimation of the drift $a(\,\cdot\,)$ and diffusion coefficient $b^2(\,\cdot\,)$ in (3.2) from high-frequency observations of (I_t) was studied by Comte, Genon-Catalot, and Rozenholc (2009). Their results are based on earlier work on nonparametric estimation for diffusion models in Comte, Genon-Catalot, and Rozenholc (2007).

The paper is organized as follows. In Section 3.2 we present the notation used throughout, formulate our general assumption on (X_t) for the asymptotic theory, and define a tractable class of prediction-based estimating functions. Section 3.3 contains functional versions of the classic Euler approximation

$$X_{t_i^n} \,|\, \mathcal{F}_{t_{i-1}^n} \approx \mathcal{N}\left(X_{t_{i-1}^n} + \Delta_n a(X_{t_{i-1}^n};\theta), \Delta_n b^2(X_{t_{i-1}^n};\theta)\right)$$

and the similar result that

$$Y_i \,|\, \mathcal{F}_{t_{i-1}^n} \approx \mathcal{N}\left(X_{t_{i-1}^n} + \Delta_n \frac{1}{2} a(X_{t_{i-1}^n};\theta), \Delta_n \frac{1}{3} b^2(X_{t_{i-1}^n};\theta)\right)$$

for small values of Δ_n. The latter approximation was essentially pointed out by Gloter (2000). Formally, these approximations take the form of power expansions and we refer to them as *Euler-Itô expansions* in this paper. They serve as building blocks for the asymptotic results in Section 3.5. Section 3.4 is devoted to limit theorems for integrated diffusions which, whenever possible, illustrate the advantage of working directly with the integrated observations. Section 3.6 concludes. Proofs are deferred to Section 3.7 and we include any auxiliary results in Section 3.8.

3.2 Preliminaries

In this section we present the general notation used throughout the paper, formulate our main assumption on the underlying diffusion model (X_t), and define a tractable class of prediction-based estimating functions.

3.2.1 Notation

1. The true parameter is denoted by θ_0.

2. We denote the state space of (X_t) by $(S, \mathscr{B}(S))$ and allow S to be an arbitrary open interval, i.e. $S = (l, r)$ for $-\infty \le l < r \le \infty$, equipped with its Borel σ-algebra $\mathscr{B}(S)$.

3. In short, we write $\mu_\theta(f) = \int_S f(x)\mu_\theta(dx)$ for functions $f : S \to \mathbb{R}$ and denote by $\mathscr{L}^p(\mu_\theta)$ the space of equivalence classes of p-integrable functions w.r.t. the invariant measure μ_θ.

4. For any diffusion process (X_t), the *potential* is defined as the operator $f \mapsto U_\theta(f)$, where
$$U_\theta(f)(x) = \int_0^\infty P_t^\theta f(x)\,\mathrm{d}t \tag{3.7}$$
and $(P_t^\theta)_{t\geq 0}$ denotes the family of transition operators,
$$P_t^\theta f(x) = \mathbb{E}_\theta\left(f(X_t) \mid X_0 = x\right).$$

5. For random variables Y and Z defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we write $Y \leq_C Z$ if there exists a constant $C > 0$ such that $Y \leq C \cdot Z$, $\mathbb{P}$-almost surely.

To define some function spaces of particular interest, recall that $f : S \times \Theta \to \mathbb{R}$ is said to be of *polynomial growth in* x if $|f(x;\theta)| \leq_C 1 + |x|^C$ for all $x \in S$.

6. We denote by $\mathcal{C}_p^{j,k}(S \times \Theta)$, $j, k \geq 0$, the class of real-valued functions $f(x;\theta)$ such that
   - f is j times continuously differentiable w.r.t. x;
   - f is k times continuously differentiable w.r.t. $\theta_1, \ldots, \theta_d$;
   - f and all partial derivatives $\partial_x^{j_1}\partial_{\theta_1}^{k_1}\cdots\partial_{\theta_d}^{k_d} f$, $j_1 \leq j$, $k_1 + \cdots + k_d \leq k$, are of polynomial growth in x.

   Similarly, we define $\mathcal{C}_p^j(S)$.

7. We let
$$\mathscr{H}_\theta^2 = \{f \in \mathcal{C}_p^4(S) : \mu_\theta(f) = 0, U_\theta(f) \in \mathcal{C}_p^2(S)\}. \tag{3.8}$$

8. The infinitesimal generator of (X_t) is denoted by $\mathcal{A}_\theta$ and the corresponding domain by $\mathcal{D}_{\mathcal{A}_\theta}$. For this paper, it suffices to note that if (X_t) satisfies Condition 3.2.1 below, $\mathcal{C}_p^2(S) \subset \mathcal{D}_{\mathcal{A}_\theta}$ and the explicit representation $\mathcal{A}_\theta f = \mathcal{L}_\theta f$, where
$$\mathcal{L}_\theta f(x) = a(x;\theta)\partial_x f(x) + \frac{1}{2}b^2(x;\theta)\partial_x^2 f(x) \tag{3.9}$$
holds for all $f \in \mathcal{C}_p^2(S)$; see e.g. Kessler (2000).

9. Finally, for use in the appendices, $R(\Delta, x;\theta)$ denotes a generic function such that
$$|R(\Delta, x;\theta)| \leq_C F(x;\theta), \tag{3.10}$$
where F is of polynomial growth in x.

3.2.2 Model assumption

To establish asymptotic results for integrated diffusions of the general form (3.1)-(3.2), we impose the following dependence and regularity conditions on (X_t):

Condition 3.2.1. *For any $\theta \in \Theta$, the stochastic differential equation*

$$dX_t = a(X_t;\theta)dt + b(X_t;\theta)dB_t, \;\; X_0 \sim \mu_\theta$$

has a weak solution $(\Omega, (\mathcal{F}_t), \mathbb{P}_\theta, (B_t), (X_t))$ *for which*

- X_0 *is independent of* (B_t),
- $\mathcal{F}_t = \sigma\left(X_0, (B_s)_{s\leq t}\right)$,
- (X_t) *is stationary,* ρ*-mixing under* $\mathbb{P}_\theta$.

Moreover, the a priori triplet (a, b, μ_θ) *satisfies the regularity conditions*

- $a, b \in \mathcal{C}_p^{2,0}(S \times \Theta)$,
- $|a(x;\theta)| + |b(x;\theta)| \leq_C 1 + |x|$,
- $b(x;\theta) > 0$ *for* $x \in S$,
- $\int_S |x|^k \mu_\theta(dx) < \infty$ *for all* $k \geq 1$.

For the discretized filtration $\{\mathcal{F}_{t_i^n}\}$ we let $\mathcal{F}_i^n := \mathcal{F}_{t_i^n}$. In particular, we want Condition 3.2.1 to be satisfied for the true, unknown, parameter θ_0. The following restriction on Θ ensures that we obtain asymptotic normality of the prediction-based estimators $\hat{\theta}_n$ in Section 3.5 by standard arguments.

Condition 3.2.2. *Suppose that*

- $\theta_0 \in \mathrm{int}(\Theta)$;
- Θ *is convex.*

Here $\mathrm{int}(\Theta)$ denotes the interior of Θ. The notation $\mu_0 = \mu_{\theta_0}$, $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$, etc., is applied throughout the paper.

3.2.3 Prediction-based estimating functions

Prediction-based estimating functions were proposed by M. Sørensen (2000, 2011) as a versatile estimation framework for non-Markovian diffusion-type models. In this paper, we consider the class of estimating functions

$$G_n(\theta) = \sum_{i=q+1}^{n} \sum_{j=1}^{N} \pi_{i-1,j} \left[f_j(Y_i) - \breve{\pi}_{i-1,j}(\theta)\right] \tag{3.11}$$

where $\{f_j\}_{j=1}^N$ is a finite set of real-valued functions in $\mathscr{L}^2(\mu_\theta)$ and for each $j \in \{1, \ldots, N\}$, $\breve{\pi}_{i-1,j}(\theta)$ denotes the orthogonal $\mathscr{L}^2(\mu_\theta)$-projection of $f_j(Y_i)$ onto a finite-dimensional subspace

$$\mathcal{P}_{i-1,j} = \text{span}\left\{1, f_j\left(Y_{i-1}\right), \ldots, f_j\left(Y_{i-q_j}\right)\right\} \subset \mathscr{L}^2(\mu_\theta) \tag{3.12}$$

where $q_j \geq 0$. The coefficients $\pi_{i-1,j}$ that appear in (3.11) are d-dimensional column vectors with entries in $\mathcal{P}_{i-1,j}$.

The collection $\{\mathcal{P}_{i-1,j}\}_{ij}$ are known as *predictor spaces*. Hence, what we predict are values of $f_j(Y_i)$ for $i \geq q+1$ where $q := \max_{1 \leq j \leq N} q_j$. Since every predictor space $\mathcal{P}_{i-1,j}$ is closed, the $\mathscr{L}^2(\mu_\theta)$-projection of $f_j(Y_i)$ onto $\mathcal{P}_{i-1,j}$ is well-defined and uniquely determined by the normal equations

$$\mathbb{E}_\theta\left(\pi\left[f_j(Y_i) - \breve{\pi}_{i-1,j}(\theta)\right]\right) = 0 \tag{3.13}$$

for all $\pi \in \mathcal{P}_{i-1,j}$. Moreover, by restricting our attention to stationary diffusion models (X_t) and predictor spaces of the form (3.12), the orthogonal projection $\breve{\pi}_{i-1,j}(\theta) = \breve{a}_n(\theta)_j^T Z_{i-1,j}$ where

$$Z_{i-1,j} = \left(1, f_j\left(Y_{i-1}\right), \ldots, f_j\left(Y_{i-q_j}\right)\right)^T$$

and $\breve{a}_n(\theta)_j^T$ denotes the (q_j+1)-dimensional coefficient vector

$$\breve{a}_n(\theta)_j^T = \left(\breve{a}_n(\theta)_{j0}, \breve{a}_n(\theta)_{j1} \ldots, \breve{a}_n(\theta)_{jq_j}\right)$$

determined by the moment conditions

$$\mathbb{E}_\theta\left[Z_{q_j,j} f_j(Y_{q_j+1})\right] - \mathbb{E}_\theta\left[Z_{q_j,j} Z_{q_j,j}^T\right] \breve{a}_n(\theta)_j = 0. \tag{3.14}$$

Note that in the simplest case of $q_j = 0$, $\mathcal{P}_{i-1,j} = \text{span}\{1\}$ and it follows immediately from the normal equations (3.13) that $\breve{\pi}_{i-1,j}(\theta) = \mathbb{E}_\theta f_j(Y_1)$.

We obtain an estimator $\hat{\theta}_n$ by solving the estimating equation

$$G_n(\theta) = 0$$

and refer to $\hat{\theta}_n$ as a G_n*-estimator*.

Remark 3.2.3. For most diffusion models, the distribution of Y_i is difficult to determine for a fixed $\Delta_n > 0$. Therefore, the evaluation of moments in (3.14) poses a difficult problem, with no general solution. As noted by Ditlevsen and M. Sørensen (2004), a restriction to polynomial predictor functions $f_j(y) = y^{\beta_j}$, $\beta_j \in \mathbb{N}$, often enables us to find the necessary moments by integrating over mixed moments of (X_t) and leads to explicit prediction-based estimating functions for the Pearson diffusions defined in Forman and M. Sørensen (2008).

3.3 Euler-Itô expansions

This section is devoted to power expansions for functionals of diffusion and integrated diffusion processes observed over a small time interval $t_i^n - t_{i-1}^n = \Delta_n$. Essentially, the following results provide a bridge between the asymptotic theory in Chapter 2 and that of the present paper as a somewhat lengthy, but straightforward, extension. The results are formulated w.r.t. an arbitrary probability measure $\mathbb{P}_\theta$.

3.3.1 Diffusion processes

The following expansion appears in various guises in the literature on statistical inference for stochastic differential equations; see e.g. Kessler (1997).

Proposition 3.3.1. *Let $f \in \mathcal{C}_p^4(S)$. Then, there exist $\mathcal{F}_i^n$-measurable random variables $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$ such that*

$$f(X_{t_i^n}) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2} \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta)\varepsilon_{1,i} + \varepsilon_{2,i} \tag{3.15}$$

where

- $\varepsilon_{1,i} \sim \mathcal{N}(0,1)$ *and independent of* $\mathcal{F}_{i-1}^n$,
- $\varepsilon_{2,i}$ *satisfies the moment expansions*

$$\mathbb{E}_\theta\left(\varepsilon_{2,i} \mid \mathcal{F}_{i-1}^n\right) = \Delta_n \mathcal{L}_\theta f(X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta), \tag{3.16}$$

$$\mathbb{E}_\theta\left(\varepsilon_{2,i}^2 \mid \mathcal{F}_{i-1}^n\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta). \tag{3.17}$$

3.3.2 Integrated diffusions

To establish a similar result for functionals of the integrated process, we rely on earlier work by Gloter (2000) as well as k'th order Taylor expansions of the form

$$f(Y_i) = \sum_{j=0}^{k-1} \frac{1}{j!} \partial_x^j f(X_{t_{i-1}^n})(Y_i - X_{t_{i-1}^n})^j + \frac{1}{k!}\partial_x^k f(Z_i^n)(Y_i - X_{t_{i-1}^n})^k, \tag{3.18}$$

where Z_i^n denotes a random variable between $X_{t_{i-1}^n}$ and Y_i, i.e.

$$Z_i^n = X_{t_{i-1}^n} + s(Y_i - X_{t_{i-1}^n})$$

for some $s \in (0,1)$.

The following lemma provides an upper bound for the remainder term in (3.18) for a given $k \geq 1$.

Lemma 3.3.2. *Let $f : S \to \mathbb{R}$ be of polynomial growth. Then, for any $k \geq 1$,*

$$\mathbb{E}_\theta \left(|f(Z_i^n)(Y_i - X_{t_{i-1}^n})|^k \,\middle|\, \mathcal{F}_{i-1}^n \right) \leq_{C_k} \Delta_n^{k/2}(1 + |X_{t_{i-1}^n}|)^{C_k}. \tag{3.19}$$

In particular, if $f \in \mathcal{C}_p^1(S)$,

$$f(Y_i) = f(X_{t_{i-1}^n}) + \partial_x f(Z_i^n)(Y_i - X_{t_{i-1}^n})$$

and Lemma 3.3.2 implies that

$$\mathbb{E}_\theta \left(|f(Y_i) - f(X_{t_{i-1}^n})|^k \,\middle|\, \mathcal{F}_{i-1}^n \right) \leq_{C_k} \Delta_n^{k/2}(1 + |X_{t_{i-1}^n}|)^{C_k}. \tag{3.20}$$

Our main result in this section is a generalization of Proposition 2.2 in Gloter (2000). Note the strong resemblance with Proposition 3.3.1.

Proposition 3.3.3. *Let $f \in \mathcal{C}_p^4(S)$. Then, there exist $\mathcal{F}_i^n$-measurable random variables $\xi_{1,i}$ and $\xi_{2,i}$ such that*

$$f(Y_i) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2} \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n}; \theta) \xi_{1,i} + \xi_{2,i} \tag{3.21}$$

where

- $\xi_{1,i} \sim \mathcal{N}(0, 1/3)$ *and independent of* $\mathcal{F}_{i-1}^n$,
- $\xi_{2,i}$ *satisfies the moment expansions*

$$\mathbb{E}_\theta \left(\xi_{2,i} \mid \mathcal{F}_{i-1}^n \right) = \Delta_n \mathcal{H}_\theta f(X_{t_{i-1}^n}) + \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n}; \theta), \tag{3.22}$$

$$\mathbb{E}_\theta \left(\xi_{2,i}^2 \mid \mathcal{F}_{i-1}^n \right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}; \theta), \tag{3.23}$$

with

$$\mathcal{H}_\theta f(x) = \frac{1}{2} \mathcal{L}_\theta f(x) - \frac{1}{12} b^2(x; \theta) \partial_x^2 f(x). \tag{3.24}$$

In particular,

$$\mathbb{E}_\theta \left(\varepsilon_{1,i} \xi_{1,i} \right) = \frac{1}{2}, \tag{3.25}$$

where $\varepsilon_{1,i}$ denotes the variable that appears in the Euler-Itô expansion (3.15).

3.4 Limit theory for integrated diffusions

As a simple application of the Euler-Itô expansion (3.21) and the corresponding bound (3.20), we derive in this section a law of large numbers (LLN) and a central limit theorem (CLT) for a class of functionals

$$\frac{1}{n} \sum_{i=1}^n f(Y_i), \tag{3.26}$$

where $f : S \to \mathbb{R}$ satisfies appropriate regularity conditions. For the remainder of the paper, all results are formulated w.r.t. the true probability measure $\mathbb{P}_0$.

Lemma 3.4.1. *Let $f \in \mathcal{C}_p^1(S)$ and suppose that (X_t) satisfies Condition 3.2.1. Then,*

$$\frac{1}{n}\sum_{i=1}^{n} f(Y_i) \xrightarrow{\mathbb{P}_0} \mu_0(f).$$

The content of Lemma 3.4.1 appears in a slightly stronger version in Proposition 2 of Gloter (2006). Remarkably, the additional rate assumption $n\Delta_n^3 \to 0$, which was necessary to obtain a CLT for functionals of diffusion processes in Chapter 2, is again sufficient to ensure convergence in law towards the *same* Gaussian limit.

Lemma 3.4.2. *Let $f \in \mathscr{H}_0^2$ and assume that (X_t) satisfies Condition 3.2.1. If $n\Delta_n^3 \to 0$, then*

$$\sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^{n} f(Y_i)\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mathcal{V}_0(f)\right),$$

where

$$\mathcal{V}_0(f) = \mu_0\left([\partial_x U_0(f) b(\,\cdot\,;\theta_0)]^2\right) = 2\mu_0\left(f U_0(f)\right). \tag{3.27}$$

Remark 3.4.3. The asymptotic variance (3.27) in Lemma 3.4.2 involves the potential $U_0(f)$, which is characterized as the solution g of the Poisson-type differential equation

$$\mathcal{L}_0(g) = -f,$$

where $\mathcal{L}_0$ denotes the differential operator (3.9) corresponding to the generator of (X_t). We refer to Section 2.3.2 for a detailed discussion of potential operators for ρ-mixing diffusion models.

3.5 Asymptotic theory

This section contains our main asymptotic results for G_n-estimators of the prediction-based estimating functions described in Section 3.2. The proofs are based on general asymptotic theory for estimating functions in M. Sørensen (2012). We confine the discussion to estimating functions of the form (3.11) where $N = 1$ and, for simplicity, write

$$G_n(\theta) = \sum_{i=q+1}^{n} \pi_{i-1}\left[f(Y_i) - \breve{\pi}_{i-1}(\theta)\right], \tag{3.28}$$

$\{\mathcal{P}_{i-1}\}_i$ for the corresponding predictor spaces, etc. The extension to estimating functions with multiple predictor functions $\{f_j\}_{j=1}^N$ was briefly discussed in Section 2.4.3.

3.5.1 Simple predictor spaces

The simplest class of estimating functions of the form (3.28) occurs for $q = 0$. In this case, the orthogonal projection $\breve{\pi}_{i-1}(\theta) = \mathbb{E}_\theta f(Y_1)$ and the one-dimensional predictor space $\mathcal{P}_{i-1}$ allows us to estimate a (sub)-parameter $\theta \in \Theta \subset \mathbb{R}$. Consistently, we suppose that $d = 1$ in the following and consider the estimating function

$$G_n(\theta) = \sum_{i=1}^{n} \left[f(Y_i) - \mathbb{E}_\theta f(Y_1) \right]. \tag{3.29}$$

The basic principle that enables us to study the asymptotic properties of G_n-estimators is to expand G_n into powers of Δ_n. In the simple case that we consider here, such an expansion follows easily from Proposition 3.3.3 since

$$f(Y_1) = f(X_0) + \Delta_n^{1/2} \partial_x f(X_0) b(X_0; \theta) \xi_{1,1} + \xi_{2,1}$$

for any $f \in \mathcal{C}_p^4(S)$ and, in particular,

$$\mathbb{E}_\theta f(Y_1) = \mu_\theta(f) + \mathbb{E}_\theta(\xi_{2,1}) = \mu_\theta(f) + \Delta_n R(\Delta_n; \theta), \tag{3.30}$$

where $|R(\Delta_n; \theta)| \leq C(\theta) < \infty$.

For the asymptotic theory we impose the following regularity conditions on G_n:

Condition 3.5.1. *Suppose that*

- $f \in \mathcal{C}_p^4(S)$,
- $f^*(x) := f(x) - \mu_0(f) \in \mathscr{H}_0^2$,
- $\theta \mapsto \mu_\theta(f) \in \mathcal{C}^1$,
- *For a connected neighbourhood $\mathcal{M}$ of θ_0 and Δ_n sufficiently small,*

$$\sup_{\theta \in \mathcal{M}} |\partial_\theta R(\Delta_n; \theta)| \leq C(\mathcal{M}). \tag{3.31}$$

For estimating functions that satisfy Condition 3.5.1, existence of a consistent, asymptotically normal G_n-estimator $\hat{\theta}_n$ holds under mild identifiability conditions:

Theorem 3.5.2. *Assume Condition 3.5.1 and suppose that $\partial_\theta \mu_\theta(f) \neq 0$ and that the identifiability condition*

$$(\mu_0 - \mu_\theta)(f) \neq 0$$

holds for all $\theta \neq \theta_0$.

- *There exists a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ which, as $n \to \infty$, is unique in any compact subset $\mathcal{K} \subset \Theta$ containing θ_0 with $\mathbb{P}_0$-probability approaching one.*
- *If, moreover, $n\Delta_n^3 \to 0$, then*

$$\sqrt{n\Delta_n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \left[\partial_\theta \mu_0(f)\right]^{-2} V_0(f)\right), \tag{3.32}$$

where

$$V_0(f) = 2\int_S f^*(x) U_0(f^*)(x) \mu_0(dx).$$

Remark 3.5.3. The identifiability conditions in Theorem 3.5.2 coincide with the findings in Chapter 2. In particular, we obtain the same Gaussian limit distribution, which enables us to re-apply the Monte Carlo algorithm proposed in Section 2.5.1 to estimate the asymptotic variance in (3.32).

3.5.2 1-lag predictor spaces

Our main result shows that for $q = 1$, prediction-based G_n-estimators remain consistent and asymptotically normal under appropriate regularity conditions.

For $q = 1$, the basis vector $Z_{i-1} = (1, f(Y_{i-1}))^T$ and the normal equations (3.14) take the form

$$\mathbb{E}_\theta\left(\begin{pmatrix} 1 \\ f(Y_1) \end{pmatrix} f(Y_2)\right) - \mathbb{E}_\theta\begin{pmatrix} 1 & f(Y_1) \\ f(Y_1) & f^2(Y_1) \end{pmatrix}\begin{pmatrix} \breve{a}_n(\theta)_0 \\ \breve{a}_n(\theta)_1 \end{pmatrix} = 0.$$

As a consequence,

$$\breve{\pi}_{i-1}(\theta) = \breve{a}_n(\theta)_0 + \breve{a}_n(\theta)_1 f(Y_{i-1}),$$

where $\breve{a}_n(\theta)_0$ and $\breve{a}_n(\theta)_1$ are uniquely determined by the moment conditions

$$\breve{a}_n(\theta)_0 = \mathbb{E}_\theta f(Y_1)\left(1 - \breve{a}_n(\theta)_1\right), \tag{3.33}$$

$$\breve{a}_n(\theta)_1 = \frac{\mathbb{E}_\theta\left[f(Y_1)f(Y_2)\right] - \left[\mathbb{E}_\theta f(Y_1)\right]^2}{\mathbb{V}ar_\theta f(Y_1)}. \tag{3.34}$$

Consistent with a two-dimensional predictor space $\mathcal{P}_{i-1}$ we suppose that $d = 2$ in the following and consider the estimating function

$$G_n(\theta) = \sum_{i=2}^n \begin{pmatrix} 1 \\ f(Y_{i-1}) \end{pmatrix}\left[f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1})\right] \tag{3.35}$$

where, unlike in Section 3.5.1, the expansion of G_n into powers of Δ_n poses a difficult problem.

Based on the Euler-Itô expansions of Section 3.3, we start by expanding the projection coefficients $\breve{a}_n(\theta)_0$ and $\breve{a}_n(\theta)_1$ in (3.33)-(3.34). As the proof is a bit long, we formulate the result as a separate lemma.

Lemma 3.5.4. *For $f \in \mathcal{C}_p^4(S)$, the projection coefficient $\breve{a}_n(\theta) = (\breve{a}_n(\theta)_0, \breve{a}_n(\theta)_1)^T$ has a power expansion*

$$\breve{a}_n(\theta) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Delta_n \begin{pmatrix} -K_f(\theta)\mu_\theta(f) \\ K_f(\theta) \end{pmatrix} + \Delta_n^{3/2} R(\Delta_n;\theta) \tag{3.36}$$

where $|R(\Delta_n;\theta)| \leq C(\theta)$ and the constant

$$K_f(\theta) = \mathbb{V}ar_\theta f(X_0)^{-1} \left[\mu_\theta(f\mathcal{L}_\theta f) + \frac{1}{6}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right]. \tag{3.37}$$

Recall that a similar power expansion of $\breve{a}_n(\theta)$ was shown for discretely observed diffusion models in Chapter 2 which, for comparison, differs mainly by *not* having the additional term

$$\mathbb{V}ar_\theta f(X_0)^{-1}\frac{1}{6}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)$$

in the expression for $K_f(\theta)$.

Condition 3.5.5. *Suppose that*

- $f \in \mathcal{C}_p^4(S)$,
- $f_1^*(x) = K_f(\theta_0)\left[\mu_0(f) - f(x)\right] \in \mathscr{H}_0^2$,
- $f_2^*(x) = f(x)\mathcal{L}_0 f(x) + \frac{1}{6}[b(x;\theta_0)\partial_x f(x)]^2 - K_f(\theta_0)f(x)\left[f(x) - \mu_0(f)\right] \in \mathscr{H}_0^2$,
- $\theta \mapsto \mu_\theta(f) \in \mathcal{C}^1$, $\theta \mapsto K_f(\theta) \in \mathcal{C}^1$ *and*

$$\sup_{\theta\in\mathcal{M}} \|\partial_{\theta^T} R(\Delta_n;\theta)\| \leq C(\mathcal{M}), \tag{3.38}$$

for a compact, convex subspace $\mathcal{M}$ containing θ_0 and Δ_n sufficiently small.

The matrix norm $\|\cdot\|$ in (3.38) can be chosen arbitrarily and we suppose for convenience that $\|\cdot\|$ is submultiplicative. The following lemma ensures the existence of a consistent sequence of G_n-estimators in Theorem 3.5.7.

Lemma 3.5.6. *Assume that Condition 3.5.5 holds. Then, for any $\theta \in \Theta$,*

$$(n\Delta_n)^{-1}G_n(\theta) \xrightarrow{\mathbb{P}_0} \gamma(\theta_0;\theta)$$

where

$$\gamma(\theta_0;\theta) = \begin{pmatrix} K_f(\theta)(\mu_\theta - \mu_0)(f) \\ \mu_0(f\mathcal{L}_0 f) + \frac{1}{6}\mu_0\left([b(\,\cdot\,;\theta_0)\partial_x f]^2\right) - K_f(\theta)\left[\mu_0(f^2) - \mu_0(f)\mu_\theta(f)\right] \end{pmatrix}. \tag{3.39}$$

Moreover,

$$\sup_{\theta\in\mathcal{M}} \left\|(n\Delta_n)^{-1}\partial_{\theta^T}G_n(\theta) - W(\theta)\right\| \xrightarrow{\mathbb{P}_0} 0 \tag{3.40}$$

where the uniform limit

$$W(\theta) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix} \begin{pmatrix} \partial_{\theta_1}\left[K_f(\theta)\mu_\theta(f)\right] & \partial_{\theta_2}\left[K_f(\theta)\mu_\theta(f)\right] \\ -\partial_{\theta_1}K_f(\theta) & -\partial_{\theta_2}K_f(\theta) \end{pmatrix}.$$

Theorem 3.5.7. *Assume Condition 3.5.5 and suppose that $W(\theta)$ is non-singular and that the identifiability condition*

$$\gamma(\theta_0;\theta) \neq 0$$

holds for all $\theta \neq \theta_0$.

- *There exists a consistent sequence of G_n-estimators $(\hat\theta_n)$ which, as $n\to\infty$, is unique in any compact subset $\mathcal{K}\subset\Theta$ containing θ_0 with $\mathbb{P}_0$-probability approaching one.*
- *If, moreover, $n\Delta_n^2 \to 0$, then*

$$\sqrt{n\Delta_n}\left(\hat\theta_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2\left(0, \left[W(\theta_0)^{-1}V_0(f)(W(\theta_0)^{-1})^T\right]\right), \tag{3.41}$$

where

$$V_0(f) = \begin{pmatrix} \mu_0\left([\partial_x U_0(f_1^*)b(\,\cdot\,;\theta_0)]^2\right) & \mathbb{C}ov(f) \\ \mathbb{C}ov(f) & \mu_0\left([\partial_x U_0(f_2^*) + f\partial_x f]^2\, b^2(\,\cdot\,;\theta_0)\right) \end{pmatrix},$$

with

$$\mathbb{C}ov(f) = \mu_0\left(\partial_x U_0(f_1^*)\left[\partial_x U_0(f_2^*) + f\partial_x f\right] b^2(\,\cdot\,;\theta_0)\right).$$

Remark 3.5.8. Compared to the findings in Chapter 2, the lower order $O(\Delta_n^{3/2})$ of the remainder term in the expansion (3.36) of $\breve{a}_n(\theta)$ forces us to assume $n\Delta_n^2 \to 0$. The same assumption appears in Gloter (2006) to ensure asymptotic normality for a class of minimum contrast estimators with integrated observations.

3.6 Extensions and concluding remarks

For integrated diffusions observed on $[0,1]$, Gloter and Gobet (2008) prove that the statistical model satisfies the LAMN property and that the optimal rate of *any* estimator of a parameter in the diffusion coefficient $b^2(\,\cdot\,;\theta)$ is $\sqrt{n}$. The ergodic scenario of this paper has yet to be considered, however, as we do not distinguish between drift and diffusion parameters in (3.2), the $\sqrt{n\Delta_n}$ rate of the previous section is all we could hope for. In comparison, the minimum contrast estimators in Gloter (2006) attain a rate of $\sqrt{n\Delta_n}$ for parameters in the drift and $\sqrt{n}$ for diffusion parameters, similar to the efficient estimators in M. Sørensen (2017).

One extension of interest would be to include a jump component into the dynamics of (X_t), corresponding to jumps in volatility. This extension has the particular feature that jumps in (X_t) lead to changes in the trend of (I_t) and *not* to path discontinuities. As a consequence, threshold estimators developed for processes with jumps observed at high frequency (see e.g. Mancini (2009)) are not directly transferable. The first to propose a general test for the presence of volatility jumps using change-point theory are Bibinger, Jirak, and Vetter (2017). How and whether the same principle can be applied for parametric inference is an interesting topic of further research.

3.7 Appendix A: Proofs

Proof of Proposition 3.3.1. By Itô's formula,

$$f(X_{t_i^n}) = f(X_{t_{i-1}^n}) + \int_{(i-1)\Delta_n}^{i\Delta_n} \mathcal{L}_\theta f(X_s)\,\mathrm{d}s + \int_{(i-1)\Delta_n}^{i\Delta_n} \partial_x f(X_s) b(X_s;\theta) dB_s$$

and letting

$$\begin{aligned}
\varepsilon_{1,i} &= \Delta_n^{-1/2} \int_{(i-1)\Delta_n}^{i\Delta_n} dB_s, \qquad (3.42)\\
A_i &= \int_{(i-1)\Delta_n}^{i\Delta_n} \mathcal{L}_\theta f(X_s)\,\mathrm{d}s,\\
B_i &= \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\partial_x f(X_s) b(X_s;\theta) - \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta)\right] dB_s,\\
\varepsilon_{2,i} &= A_i + B_i,
\end{aligned}$$

we obtain an expansion of the form

$$f(X_{t_i^n}) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2} \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta)\varepsilon_{1,i} + \varepsilon_{2,i}$$

where, clearly, $\varepsilon_{1,i}$ and $\varepsilon_{2,i}$ are $\mathcal{F}_i^n$-measurable and $\varepsilon_{1,i} \sim \mathcal{N}(0,1)$ and independent of $\mathcal{F}_{i-1}^n$.

The conditional moment expansions (3.16)-(3.17) of $\varepsilon_{2,i}$ require a bit more effort; by applying Fubini's theorem followed by Lemma 2.10.2,

$$\begin{aligned}
\mathbb{E}_\theta\left(A_i \,\middle|\, \mathcal{F}^n_{i-1}\right) &= \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}_\theta\left(\mathcal{L}_\theta f(X_s) \,\middle|\, \mathcal{F}^n_{i-1}\right) \mathrm{d}s \\
&= \int_0^{\Delta_n} \mathbb{E}_\theta\left(\mathcal{L}_\theta f(X_{t^n_{i-1}+u}) \,\middle|\, \mathcal{F}^n_{i-1}\right) \mathrm{d}u \\
&= \int_0^{\Delta_n} \left[\mathcal{L}_\theta f(X_{t^n_{i-1}}) + u \cdot R(u, X_{t^n_{i-1}};\theta)\right] \mathrm{d}u \\
&= \Delta_n \mathcal{L}_\theta f(X_{t^n_{i-1}}) + \Delta_n^2 R(\Delta_n, X_{t^n_{i-1}};\theta).
\end{aligned}$$

Moreover, since $\mathbb{E}_\theta\left(\int_0^t [\partial_x f(X_s) b(X_s;\theta)]^2 \, \mathrm{d}s\right) = \mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right) \cdot t < \infty$, the stochastic integral $\int_0^t \partial_x f(X_s) b(X_s;\theta) dB_s$ constitutes a true $\mathbb{P}_\theta$-martingale on $[0,\infty)$ and, in turn,

$$M_t := \int_0^t \left[\partial_x f(X_s) b(X_s;\theta) - \partial_x f(X_{t^n_{i-1}}) b(X_{t^n_{i-1}};\theta)\right] dB_s$$

satisfies the martingale property $\mathbb{E}_\theta\left(M_t \mid \mathcal{F}_s\right) = M_s$ for $t \geq s \geq t^n_{i-1}$. In particular,

$$\mathbb{E}_\theta\left(B_i \,\middle|\, \mathcal{F}^n_{i-1}\right) = \mathbb{E}_\theta\left(M_{i\Delta_n} - M_{(i-1)\Delta_n} \,\middle|\, \mathcal{F}^n_{i-1}\right) = 0,$$

which verifies (3.16).

For moments of order $k \geq 2$, we write

$$A_i = \Delta_n \mathcal{L}_\theta f(X_{t^n_{i-1}}) + \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\mathcal{L}_\theta f(X_s) - \mathcal{L}_\theta f(X_{t^n_{i-1}})\right] \mathrm{d}s \tag{3.43}$$

and observe that, by Jensen's inequality,

$$\begin{aligned}
&\left|\int_{(i-1)\Delta_n}^{i\Delta_n} \left[\mathcal{L}_\theta f(X_s) - \mathcal{L}_\theta f(X_{t^n_{i-1}})\right] \mathrm{d}s\right|^k \\
\leq\ & \Delta_n^k \cdot \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} |\mathcal{L}_\theta f(X_s) - \mathcal{L}_\theta f(X_{t^n_{i-1}})|^k \, \mathrm{d}s \\
\leq\ & \Delta_n^k \sup_{u\in[0,\Delta_n]} |\mathcal{L}_\theta f(X_{t^n_{i-1}+u}) - \mathcal{L}_\theta f(X_{t^n_{i-1}})|^k.
\end{aligned}$$

Hence, by Lemma 2.10.1,

$$\begin{aligned}
&\mathbb{E}_\theta\left(|A_i|^k \,\middle|\, \mathcal{F}^n_{i-1}\right) \\
\leq_{C_k}\ & \Delta_n^k (1 + |X_{t^n_{i-1}}|)^{C_k} + \Delta_n^k \cdot \mathbb{E}_\theta\left(\sup_{u\in[0,\Delta_n]} |\mathcal{L}_\theta f(X_{t^n_{i-1}+u}) - \mathcal{L}_\theta f(X_{t^n_{i-1}})|^k \,\middle|\, \mathcal{F}^n_{i-1}\right) \\
\leq_{C_k}\ & \Delta_n^k (1 + |X_{t^n_{i-1}}|)^{C_k}
\end{aligned}$$

or, *equivalently*,

$$\mathbb{E}_\theta\left(|A_i|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) = \Delta_n^k R(\Delta_n, X_{t_{i-1}^n};\theta).$$

Similarly, if we let $h(x;\theta) = \partial_x f(x) b(x;\theta)$, the Burkholder-Davis-Gundy inequality (see e.g. Jacod and Protter (2012)), Jensen's inequality and Lemma 2.10.1 imply that for all $k \geq 2$,

$$\begin{aligned}
& \mathbb{E}_\theta\left(|B_i|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
= \quad & \mathbb{E}_\theta\left(\left|\int_{(i-1)\Delta_n}^{i\Delta_n} \left[h(X_s;\theta) - h(X_{t_{i-1}^n};\theta)\right] dB_s\right|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
\leq_{C_k} \quad & \mathbb{E}_\theta\left(\left[\int_{(i-1)\Delta_n}^{i\Delta_n} \left[h(X_s;\theta) - h(X_{t_{i-1}^n};\theta)\right]^2 \mathrm{d}s\right]^{k/2} \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
\leq \quad & \Delta_n^{k/2} \cdot \mathbb{E}_\theta\left(\Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} |h(X_s;\theta) - h(X_{t_{i-1}^n};\theta)|^k \,\mathrm{d}s \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
\leq \quad & \Delta_n^{k/2} \cdot \mathbb{E}_\theta\left(\sup_{u\in[0,\Delta_n]} |h(X_{t_{i-1}^n+u};\theta) - h(X_{t_{i-1}^n};\theta)|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
\leq_{C_k} \quad & \Delta_n^k (1 + |X_{t_{i-1}^n}|)^{C_k}
\end{aligned}$$

and since $|\varepsilon_{2,i}|^k \leq_{C_k} |A_i|^k + |B_i|^k$,

$$\mathbb{E}_\theta\left(|\varepsilon_{2,i}|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) = \Delta_n^k R(\Delta_n, X_{t_{i-1}^n};\theta)$$

with $k = 2$ as a special case. □

Proof of Lemma 3.3.2. By Jensen's inequality,

$$|Y_i - X_{t_{i-1}^n}|^k \leq \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} |X_s - X_{t_{i-1}^n}|^k \,\mathrm{d}s \leq \sup_{u\in[0,\Delta_n]} |X_{t_{i-1}^n+u} - X_{t_{i-1}^n}|^k$$

and, hence, Lemma 2.10.1 implies that

$$\mathbb{E}_\theta\left(|Y_i - X_{t_{i-1}^n}|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \leq_{C_k} \Delta_n^{k/2}(1 + |X_{t_{i-1}^n}|)^{C_k}. \tag{3.44}$$

Moreover, since f is of polynomial growth,

$$|f(Z_i^n)|^k \leq_{C_k} 1 + |X_{t_{i-1}^n}|^{C_k} + |Y_i - X_{t_{i-1}^n}|^{C_k}$$

and the result follows by a double application of (3.44). □

Proof of Proposition 3.3.3. We start by proving the result for the identity map $f(x) = x$. In this case, $\partial_x f \equiv 1$, $\partial_x^2 f \equiv 0$ and the Euler-Itô expansion (3.21) takes the form

$$Y_i = X_{t_{i-1}^n} + \Delta_n^{1/2} b(X_{t_{i-1}^n};\theta)\xi_{1,i}^* + \xi_{2,i}^*, \tag{3.45}$$

where asterisks ($*$) have been added to distinguish the remainder terms from the general case. In particular, we must have

$$\mathbb{E}_\theta\left(\xi_{2,i}^* \,\middle|\, \mathcal{F}_{i-1}^n\right) = \Delta_n \frac{1}{2} a(X_{t_{i-1}^n};\theta) + \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n};\theta). \tag{3.46}$$

By applying the auxiliary Lemma 3.8.1 to the stochastic integral,

$$\begin{aligned} Y_i - X_{t_{i-1}^n} &= \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \left(\int_{(i-1)\Delta_n}^{s} a(X_u;\theta)\,\mathrm{d}u + \int_{(i-1)\Delta_n}^{s} b(X_u;\theta) dB_u \right) \mathrm{d}s \\ &= \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_{(i-1)\Delta_n}^{s} a(X_u;\theta)\,\mathrm{d}u + \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} (i\Delta_n - s) b(X_s;\theta) dB_s, \end{aligned}$$

and, in turn, this yields an expansion of the form (3.45) by letting

$$\begin{aligned} \xi_{1,i}^* &= \Delta_n^{-3/2} \int_{(i-1)\Delta_n}^{i\Delta_n} (i\Delta_n - s) dB_s, \\ A_i &= \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_{(i-1)\Delta_n}^{s} a(X_u;\theta)\,\mathrm{d}u, \\ B_i &= \Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} \left[b(X_s;\theta) - b(X_{t_{i-1}^n};\theta) \right] (i\Delta_n - s) dB_s, \\ \xi_{2,i}^* &= A_i + B_i. \end{aligned}$$

To verify the properties of $\xi_{1,i}^*$ and $\xi_{2,i}^*$, respectively, we observe directly that both variables are measurable w.r.t. $\mathcal{F}_i^n$, $\xi_{1,i}^*$ is Gaussian and independent of $\mathcal{F}_{i-1}^n$ and, by the martingale property of $\int_0^t (i\Delta_n - s) dB_s$, $\mathbb{E}_\theta(\xi_{1,i}^*) = 0$. Moreover, by Itô's isometry

$$\mathbb{E}_\theta((\xi_{1,i}^*)^2) = \Delta_n^{-3} \int_{(i-1)\Delta_n}^{i\Delta_n} (i\Delta_n - s)^2\,\mathrm{d}s = \frac{1}{3}.$$

As in the proof of Proposition 3.3.1, the conditional moment expansions of $\xi_{2,i}^*$ require some effort. From the martingale property of $\int_0^t b(X_s;\theta)(i\Delta_n - s)dB_s$, we conclude that

$$\begin{aligned} &\mathbb{E}_\theta\left(B_i \,\middle|\, \mathcal{F}_{i-1}^n\right) \\ =\;& \Delta_n^{-1} \cdot \mathbb{E}_\theta\left(\int_{(i-1)\Delta_n}^{i\Delta_n} \left[b(X_s;\theta) - b(X_{t_{i-1}^n};\theta) \right] (i\Delta_n - s) dB_s \,\middle|\, \mathcal{F}_{i-1}^n \right) \\ =\;& \Delta_n^{-1} \cdot \mathbb{E}_\theta\left(\int_{(i-1)\Delta_n}^{i\Delta_n} b(X_s;\theta)(i\Delta_n - s) dB_s \,\middle|\, \mathcal{F}_{i-1}^n \right) - \Delta_n^{1/2} b(X_{t_{i-1}^n};\theta) \mathbb{E}_\theta\left(\xi_{1,i}^* \,\middle|\, \mathcal{F}_{i-1}^n\right) \\ =\;& 0. \end{aligned}$$

Therefore, $\mathbb{E}_\theta(\xi^*_{2,i} \,|\, \mathcal{F}^n_{i-1}) = \mathbb{E}_\theta(A_i \,|\, \mathcal{F}^n_{i-1})$ and a double application of Fubini's theorem followed by Lemma 2.10.2 shows that

$$\begin{aligned}
\mathbb{E}_\theta\left(A_i \,\middle|\, \mathcal{F}^n_{i-1}\right) &= \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_{(i-1)\Delta_n}^{s} \mathbb{E}_\theta\left(a(X_u;\theta) \,\middle|\, \mathcal{F}^n_{i-1}\right) \mathrm{d}u \\
&= \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_0^{s-t^n_{i-1}} \mathbb{E}_\theta\left(a(X_{t^n_{i-1}+v};\theta) \,\middle|\, \mathcal{F}^n_{i-1}\right) \mathrm{d}v \\
&= \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_0^{s-t^n_{i-1}} \left[a(X_{t^n_{i-1}};\theta) + vR(v, X_{t^n_{i-1}};\theta)\right] \mathrm{d}v \\
&= \Delta_n \frac{1}{2} a(X_{t^n_{i-1}};\theta) + \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_0^{s-t^n_{i-1}} vR(v, X_{t^n_{i-1}};\theta)\,\mathrm{d}v
\end{aligned}$$

and the moment expansion (3.46) follows since

$$\Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \int_0^{s-t^n_{i-1}} vR(v, X_{t^n_{i-1}};\theta)\,\mathrm{d}v = \Delta_n^2 R(\Delta_n, X_{t^n_{i-1}};\theta)$$

by direct verification.[2]

To show that $\mathbb{E}_\theta((\xi^*_{2,i})^2 \,|\, \mathcal{F}^n_{i-1}) = \Delta_n^2 R(\Delta_n, X_{t^n_{i-1}};\theta)$, we apply that

$$|A_i|^k \le \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} \mathrm{d}s \left|\int_{(i-1)\Delta_n}^{s} a(X_u;\theta)\,\mathrm{d}u\right|^k \le \sup_{s\in[0,\Delta_n]} \left|\int_{t^n_{i-1}}^{t^n_{i-1}+s} a(X_u;\theta)\,\mathrm{d}u\right|^k$$

for all $k \ge 1$ by Jensen's inequality. Moreover, for any $t \ge 0$,

$$\begin{aligned}
&\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]} \left|\int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u\right|^k \,\middle|\, \mathcal{F}_t\right) \\
\le\ &\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]} s^{k-1}\int_t^{t+s} |a(X_u;\theta)|^k\,\mathrm{d}u \,\middle|\, \mathcal{F}_t\right) \\
=\ &\mathbb{E}_\theta\left(\Delta_n^{k-1}\int_t^{t+\Delta_n} |a(X_u;\theta)|^k\,\mathrm{d}u \,\middle|\, \mathcal{F}_t\right)
\end{aligned}$$

and by the linear growth assumption of $a(\,\cdot\,;\theta)$ in Condition 3.2.1,

$$|a(X_u;\theta)|^k \le_{C_k} 1 + |X_t|^k + |X_u - X_t|^k$$

[2] In fact, we see that the higher order expansion $\mathbb{E}_\theta(\xi^*_{2,i} \,|\, \mathcal{F}^n_{i-1}) = \Delta_n \frac{1}{2} a(X_{t^n_{i-1}};\theta) + \Delta_n^2 R(\Delta_n, X_{t^n_{i-1}};\theta)$ holds. The use of $\Delta_n^{3/2} R(\Delta_n, X_{t^n_{i-1}};\theta)$ will be clear when we generalize to arbitrary $f \in \mathcal{C}^4_p(S)$.

which, in turn, shows that

$$
\begin{aligned}
& \mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]}\left|\int_t^{t+s} a(X_u;\theta)\,\mathrm{d}u\right|^k \,\middle|\, \mathcal{F}_t\right) \\
\leq_{C_k} \quad & \Delta_n^k(1+|X_t|^k) + \Delta_n^{k-1}\int_t^{t+\Delta_n} \mathbb{E}_\theta\left(|X_u - X_t|^k \,\middle|\, \mathcal{F}_t\right)\mathrm{d}u \\
\leq_{C_k} \quad & \Delta_n^k(1+|X_t|)^{C_k} + \Delta_n^k\cdot\mathbb{E}_\theta\left(\sup_{v\in[0,\Delta_n]}|X_{t+v}-X_t|^k \,\middle|\, \mathcal{F}_t\right) \\
\leq_{C_k} \quad & \Delta_n^k(1+|X_t|)^{C_k},
\end{aligned}
$$

where the final inequality follows from Lemma 2.10.1. In particular,

$$
\mathbb{E}_\theta\left(|A_i|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \leq \mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]}\left|\int_{t_{i-1}^n}^{t_{i-1}^n+s} a(X_u;\theta)\,\mathrm{d}u\right|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \leq_{C_k} \Delta_n^k(1+|X_{t_{i-1}^n}|)^{C_k}.
$$

To obtain a similar bound for $|B_i|^k$, we apply that the integral process

$$
M_t = \int_0^t \left[b(X_s;\theta) - b(X_{t_{i-1}^n};\theta)\right](i\Delta_n - s)dB_s
$$

satisfies the martingale property $\mathbb{E}_\theta\left(M_t \mid \mathcal{F}_s\right) = M_s$ for $t \geq s \geq t_{i-1}^n$. Hence, by the Burkholder-Davis-Gundy inequality, Jensen's inequality and Lemma 2.10.1,

$$
\begin{aligned}
\mathbb{E}_\theta\left(|B_i|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \quad = \quad & \mathbb{E}_\theta\left(\Delta_n^{-k}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\left[b(X_s;\theta) - b(X_{t_{i-1}^n};\theta)\right](i\Delta_n - s)dB_s\right|^k \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
\leq_{C_k} \quad & \mathbb{E}_\theta\left(\left[\int_{(i-1)\Delta_n}^{i\Delta_n}\left[b(X_s;\theta) - b(X_{t_{i-1}^n};\theta)\right]^2\mathrm{d}s\right]^{k/2} \,\middle|\, \mathcal{F}_{i-1}^n\right) \\
\leq_{C_k} \quad & \Delta_n^k(1+|X_{t_{i-1}^n}|)^{C_k}
\end{aligned}
$$

for all $k \geq 2$ and, as a consequence,

$$
\mathbb{E}_\theta\left((\xi_{2,i}^*)^2 \,\middle|\, \mathcal{F}_{i-1}^n\right) \leq_C \Delta_n^2(1+|X_{t_{i-1}^n}|)^C.
$$

The extension to arbitrary $f \in \mathcal{C}_p^4(S)$ is based on Taylor expansions of the general form (3.18): firstly, a third order Taylor expansion combined with the Euler-Itô expansion

$$
Y_i - X_{t_{i-1}^n} = \Delta_n^{1/2} b(X_{t_{i-1}^n};\theta)\xi_{1,i}^* + \xi_{2,i}^*
$$

derived above, implies that

$$
\begin{aligned}
f(Y_i) \quad &= \quad \sum_{j=0}^{2}\frac{1}{j!}\partial_x^j f(X_{t_{i-1}^n})(Y_i - X_{t_{i-1}^n})^j + \frac{1}{6}\partial_x^3 f(Z_i^n)(Y_i - X_{t_{i-1}^n})^3 \\
&= \quad f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\partial_x f(X_{t_{i-1}^n})b(X_{t_{i-1}^n};\theta)\xi_{1,i} + \xi_{2,i}
\end{aligned}
$$

where

$$\xi_{1,i} = \xi_{1,i}^* = \Delta_n^{-3/2} \int_{(i-1)\Delta_n}^{i\Delta_n} (i\Delta_n - s) dB_s \tag{3.47}$$

and the remainder term $\xi_{2,i} = \sum_{k=1}^5 \xi_{2,i}^{(k)}$ for

$$\begin{aligned}
\xi_{2,i}^{(1)} &= \partial_x f(X_{t_{i-1}^n}) \xi_{2,i}^*, \\
\xi_{2,i}^{(2)} &= \Delta_n \frac{1}{2} \partial_x^2 f(X_{t_{i-1}^n}) b^2(X_{t_{i-1}^n};\theta)(\xi_{1,i}^*)^2, \\
\xi_{2,i}^{(3)} &= \frac{1}{2} \partial_x^2 f(X_{t_{i-1}^n})(\xi_{2,i}^*)^2, \\
\xi_{2,i}^{(4)} &= \Delta_n^{1/2} \partial_x^2 f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta) \xi_{1,i}^* \xi_{2,i}^*, \\
\xi_{2,i}^{(5)} &= \frac{1}{6} \partial_x^3 f(Z_i^n)(Y_i - X_{t_{i-1}^n})^3.
\end{aligned}$$

That $\xi_{2,i}$ is measurable w.r.t. $\mathcal{F}_i^n$ follows by measurability of each constituent, in particular that of $Z_i^n = X_{t_{i-1}^n} + s(Y_i - X_{t_{i-1}^n})$, and therefore it only remains to show that $\xi_{2,i}$ satisfies the moment expansions (3.22) and (3.23), respectively.

By applying the conditional moment expansions

$$\begin{aligned}
\mathbb{E}_\theta \left(\xi_{1,i}^* \;\middle|\; \mathcal{F}_{i-1}^n \right) &= 0, \\
\mathbb{E}_\theta \left((\xi_{1,i}^*)^2 \;\middle|\; \mathcal{F}_{i-1}^n \right) &= \frac{1}{3}, \\
\mathbb{E}_\theta \left(\xi_{2,i}^* \;\middle|\; \mathcal{F}_{i-1}^n \right) &= \Delta_n \frac{1}{2} a(X_{t_{i-1}^n};\theta) + \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n};\theta), \\
\mathbb{E}_\theta \left((\xi_{2,i}^*)^2 \;\middle|\; \mathcal{F}_{i-1}^n \right) &= \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta),
\end{aligned} \tag{3.48}$$

it follows immediately that

$$\begin{aligned}
\mathbb{E}_\theta \left(\xi_{2,i}^{(1)} \;\middle|\; \mathcal{F}_{i-1}^n \right) &= \Delta_n \frac{1}{2} a(X_{t_{i-1}^n};\theta) \partial_x f(X_{t_{i-1}^n}) + \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n};\theta), \\
\mathbb{E}_\theta \left(\xi_{2,i}^{(2)} \;\middle|\; \mathcal{F}_{i-1}^n \right) &= \Delta_n \frac{1}{6} \partial_x^2 f(X_{t_{i-1}^n}) b^2(X_{t_{i-1}^n};\theta), \\
\mathbb{E}_\theta \left(\xi_{2,i}^{(3)} \;\middle|\; \mathcal{F}_{i-1}^n \right) &= \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta).
\end{aligned}$$

Furthermore, by Hölder's inequality,

$$|\mathbb{E}_\theta \left(\xi_{1,i}^* \xi_{2,i}^* \mid \mathcal{F}_{i-1}^n \right)| \leq \mathbb{E}_\theta \left((\xi_{1,i}^*)^2 \mid \mathcal{F}_{i-1}^n \right)^{1/2} \mathbb{E}_\theta \left((\xi_{2,i}^*)^2 \mid \mathcal{F}_{i-1}^n \right)^{1/2} = \Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta),$$

which shows that

$$\mathbb{E}_\theta \left(\xi_{2,i}^{(4)} \;\middle|\; \mathcal{F}_{i-1}^n \right) = \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n};\theta)$$

and, by Lemma 3.3.2,

$$\mathbb{E}_\theta \left(\xi_{2,i}^{(5)} \;\middle|\; \mathcal{F}_{i-1}^n \right) = \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n};\theta).$$

Collecting our observations,

$$\begin{aligned}
&\mathbb{E}_\theta\left(\xi_{2,i}\,\middle|\,\mathcal{F}_{i-1}^n\right)\\
=\;&\sum_{k=1}^{5}\mathbb{E}_\theta\left(\xi_{2,i}^{(k)}\,\middle|\,\mathcal{F}_{i-1}^n\right)\\
=\;&\Delta_n\left(\frac{1}{2}a(X_{t_{i-1}^n};\theta)\partial_x f(X_{t_{i-1}^n})+\frac{1}{6}b^2(X_{t_{i-1}^n};\theta)\partial_x^2 f(X_{t_{i-1}^n})\right)+\Delta_n^{3/2}R(\Delta_n,X_{t_{i-1}^n};\theta)\\
=\;&\Delta_n\left(\frac{1}{2}\mathcal{L}_\theta f(X_{t_{i-1}^n})-\frac{1}{12}b^2(X_{t_{i-1}^n};\theta)\partial_x^2 f(X_{t_{i-1}^n})\right)+\Delta_n^{3/2}R(\Delta_n,X_{t_{i-1}^n};\theta)\\
=\;&\Delta_n\mathcal{H}_\theta f(X_{t_{i-1}^n})+\Delta_n^{3/2}R(\Delta_n,X_{t_{i-1}^n};\theta).
\end{aligned}$$

Similarly, to argue that $\mathbb{E}_\theta(\xi_{2,i}^2\,|\,\mathcal{F}_{i-1}^n)=\Delta_n^2 R(\Delta_n,X_{t_{i-1}^n};\theta)$, we use that

$$\begin{aligned}
f(Y_i) &= f(X_{t_{i-1}^n})+\partial_x f(X_{t_{i-1}^n})(Y_i-X_{t_{i-1}^n})+\frac{1}{2}\partial_x^2 f(Z_i^n)(Y_i-X_{t_{i-1}^n})^2\\
&= f(X_{t_{i-1}^n})+\Delta_n^{1/2}\partial_x f(X_{t_{i-1}^n})b(X_{t_{i-1}^n};\theta)\xi_{1,i}+\xi_{2,i}
\end{aligned}$$

where the remainder term $\xi_{2,i}$ takes the explicit form

$$\xi_{2,i}=\partial_x f(X_{t_{i-1}^n})\xi_{2,i}^*+\frac{1}{2}\partial_x^2 f(Z_i^n)(Y_i-X_{t_{i-1}^n})^2.$$

In particular, this implies that

$$\xi_{2,i}^2\leq_C[\partial_x f(X_{t_{i-1}^n})]^2(\xi_{2,i}^*)^2+[\partial_x^2 f(Z_i^n)]^2(Y_i-X_{t_{i-1}^n})^4$$

and, in turn,

$$\mathbb{E}_\theta\left(\xi_{2,i}^2\,\middle|\,\mathcal{F}_{i-1}^n\right)=\Delta_n^2 R(\Delta_n,X_{t_{i-1}^n};\theta)$$

by applying (3.48) and Lemma 3.3.2, respectively.

Finally, by the defining properties (3.42), (3.47) and the Itô isometry,

$$\begin{aligned}
\mathbb{E}_\theta(\varepsilon_{1,i}\xi_{1,i}) &= \mathbb{E}_\theta\left(\Delta_n^{-1/2}\int_{(i-1)\Delta_n}^{i\Delta_n}dB_s\cdot\Delta_n^{-3/2}\int_{(i-1)\Delta_n}^{i\Delta_n}(i\Delta_n-s)dB_s\right)\\
&= \Delta_n^{-2}\cdot\int_{(i-1)\Delta_n}^{i\Delta_n}(i\Delta_n-s)\,\mathrm{d}s\\
&= \frac{1}{2}.
\end{aligned}$$

□

Proof of Lemma 3.4.1. By Lemma 2.3.1, it suffices to show that

$$\frac{1}{n}\sum_{i=1}^{n}\left[f(Y_i)-f(X_{t_{i-1}^n})\right]=o_{\mathbb{P}_0}(1) \tag{3.49}$$

and applying the upper bound (3.20) for conditional expectations,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_0\left(|f(Y_i)-f(X_{t_{i-1}^n})|\;\Big|\;\mathcal{F}_{i-1}^n\right)=\Delta_n^{1/2}\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n,X_{t_{i-1}^n};\theta_0)=o_{\mathbb{P}_0}(1),$$

$$\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_0\left(|f(Y_i)-f(X_{t_{i-1}^n})|^2\;\Big|\;\mathcal{F}_{i-1}^n\right)=\Delta_n\frac{1}{n^2}\sum_{i=1}^{n}R(\Delta_n,X_{t_{i-1}^n};\theta_0)=o_{\mathbb{P}_0}(1),$$

and, at this point, the desired conclusion (3.49) follows from the useful Lemma 9 in Genon-Catalot and Jacod (1993). □

Proof of Lemma 3.4.2. Due to Proposition 2.3.4, it is sufficient to show that

$$\sqrt{n\Delta_n}\cdot\frac{1}{n}\sum_{i=1}^{n}\left[f(Y_i)-f(X_{t_{i-1}^n})\right]=o_{\mathbb{P}_0}(1),\tag{3.50}$$

which for comparison is a strengthening of (3.49). To prove the stronger statement (3.50), note that by Proposition 3.3.3,

$$\begin{aligned}\sqrt{n\Delta_n}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_0\left(f(Y_i)-f(X_{t_{i-1}^n})\;\Big|\;\mathcal{F}_{i-1}^n\right) &= \sqrt{n\Delta_n}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_0\left(\xi_{2,i}\;\big|\;\mathcal{F}_{i-1}^n\right)\\ &= \sqrt{n\Delta_n^3}\cdot\frac{1}{n}\sum_{i=1}^{n}R(\Delta_n,X_{t_{i-1}^n};\theta_0)\\ &= o_{\mathbb{P}_0}(1),\end{aligned}$$

where we apply that $n\Delta_n^3\to 0$ and once again the higher order bound (3.20) ensures that

$$\frac{\Delta_n}{n}\sum_{i=1}^{n}\mathbb{E}_0\left(|f(Y_i)-f(X_{t_{i-1}^n})|^2\;\Big|\;\mathcal{F}_{i-1}^n\right)=\frac{\Delta_n^2}{n}\sum_{i=1}^{n}R(\Delta_n,X_{t_{i-1}^n};\theta_0)=o_{\mathbb{P}_0}(1).$$

The conclusion (3.50) follows from Lemma 9 in Genon-Catalot and Jacod (1993). □

Proof of Theorem 3.5.2. By applying the first order expansion (3.30) of $\mathbb{E}_\theta f(Y_1)$ together with Lemma 3.4.1, we see that

$$\begin{aligned}H_n(\theta) &= \frac{1}{n}\sum_{i=1}^{n}\left[f(Y_i)-\mathbb{E}_\theta f(Y_1)\right]\\ &= \frac{1}{n}\sum_{i=1}^{n}\left[f(Y_i)-\mu_\theta(f)\right]+\Delta_nR(\Delta_n;\theta)\\ &\xrightarrow{\mathbb{P}_0} H(\theta)=(\mu_0-\mu_\theta)(f)\end{aligned}\tag{3.51}$$

and for all $\theta\in\mathcal{M}$,

$$\partial_\theta H_n(\theta)=-\partial_\theta\mathbb{E}_\theta f(Y_1)=-\partial_\theta\mu_\theta(f)+\Delta_n\partial_\theta R(\Delta_n;\theta)\to-\partial_\theta\mu_\theta(f).$$

Moreover, subject to Condition 3.5.1,

$$\sup_{\theta\in\mathcal{M}} |\partial_\theta H_n(\theta) + \partial_\theta \mu_\theta(f)| = \Delta_n \sup_{\theta\in\mathcal{M}} |\partial_\theta R(\Delta_n;\theta)| \leq_{C(\mathcal{M})} \Delta_n \to 0,$$

and the eventual existence of a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ now follows from Theorem 1.58 in M. Sørensen (2012). That the estimator $\hat{\theta}_n$ will be unique in any compact subset $\mathcal{K} \subset \Theta$ that contains θ_0 with $\mathbb{P}_0$-probability approaching one as $n \to \infty$ follows from Theorem 1.59 in M. Sørensen (2012) by applying continuity of $\theta \mapsto H(\theta)$ and the identifiability assumption $H(\theta) \neq 0$ for $\theta \neq \theta_0$.

To establish asymptotic normality, note that (3.51) and the additional rate assumption $n\Delta_n^3 \to 0$ ensure that

$$\begin{aligned}
\sqrt{n\Delta_n} \cdot H_n(\theta_0) &= \sqrt{n\Delta_n} \cdot \left(\frac{1}{n} \sum_{i=1}^n f^*(Y_i) \right) + \sqrt{n\Delta_n^3} R(\Delta_n;\theta_0) \\
&= \sqrt{n\Delta_n} \cdot \left(\frac{1}{n} \sum_{i=1}^n f^*(Y_i) \right) + o_{\mathbb{P}_0}(1) \\
&\xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, V_0(f)\right),
\end{aligned}$$

where convergence in law follows directly from Lemma 3.4.2 since $f^* \in \mathscr{H}_0^2$.

At this point, the result follows from the first order Taylor expansion

$$0 = \sqrt{n\Delta_n} H_n(\theta_0) + \partial_\theta H_n(\theta_n^*) \sqrt{n\Delta_n}(\hat{\theta}_n - \theta_0),$$

where θ_n^* lies between $\hat{\theta}_n$ and θ_0 by standard arguments. □

Proof of Lemma 3.5.4. We break the proof into four steps as follows:

- **Step 1:** Expand $[\mathbb{E}_\theta f(Y_1)]^2$, $\mathbb{E}_\theta f^2(Y_1)$ and $\mathbb{E}_\theta\left[f(Y_1)f(Y_2)\right]$ into powers of Δ_n,
- **Step 2:** Eliminate $\mathcal{H}_\theta$ from the expansions,
- **Step 3:** Show that $\breve{a}_n(\theta)_1 = 1 + \Delta_n K_f(\theta) + \Delta_n^{3/2} R(\Delta_n;\theta)$,
- **Step 4:** Conclude that $\breve{a}_n(\theta)_0 = -\Delta_n K_f(\theta)\mu_\theta(f) + \Delta_n^{3/2} R(\Delta_n;\theta)$.

Step 1 By direct application of Proposition 3.3.3, we find that

$$\begin{aligned}
[\mathbb{E}_\theta f(Y_1)]^2 &= \left[\mu_\theta(f) + \Delta_n \mu_\theta(\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta) \right]^2 \\
&= \mu_\theta(f)^2 + \Delta_n 2\mu_\theta(f)\mu_\theta(\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta), \qquad (3.52)
\end{aligned}$$

and, similarly,

$$\mathbb{E}_\theta f^2(Y_1) \quad = \quad \mu_\theta(f^2) + \Delta_n \mu_\theta(\mathcal{H}_\theta f^2) + \Delta_n^{3/2} R(\Delta_n;\theta). \tag{3.53}$$

The expansion of the mixed moment $\mathbb{E}_\theta[f(Y_1)f(Y_2)]$ is more elaborate and rests on our ability to expand each individual term in

$$\mathbb{E}_\theta[f(Y_1)f(Y_2)] =$$
$$\mathbb{E}_\theta\left[\left(f(X_0) + \Delta_n^{1/2}\partial_x f(X_0) b(X_0;\theta)\xi_{1,1} + \xi_{2,1}\right)\left(f(X_{\Delta_n}) + \Delta_n^{1/2}\partial_x f(X_{\Delta_n}) b(X_{\Delta_n};\theta)\xi_{1,2} + \xi_{2,2}\right)\right]$$

by utilizing the conditional moment expansions of Section 3.3. When evaluating conditional moments of the form $\mathbb{E}_\theta\left(g(X_{\Delta_n}) \mid \mathcal{F}_0\right)$, we sometimes replace the use of Proposition 3.3.1 with that of Lemma 2.10.2 as it prevents us from imposing stricter differentiability restrictions on f.

Term 1: By Proposition 3.3.1,

$$f(X_{\Delta_n}) = f(X_0) + \Delta_n^{1/2}\partial_x f(X_0) b(X_0;\theta)\varepsilon_{1,1} + \varepsilon_{2,1} \tag{3.54}$$

and applying the tower property of conditional expectations,

$$\mathbb{E}_\theta\left[f(X_0)f(X_{\Delta_n})\right] = \mu_\theta(f^2) + \Delta_n \mu_\theta(f\mathcal{L}_\theta f) + \Delta_n^2 R(\Delta_n;\theta).$$

Term 2: To argue that

$$\Delta_n^{1/2}\mathbb{E}_\theta\left[f(X_0)\partial_x f(X_{\Delta_n}) b(X_{\Delta_n};\theta)\xi_{1,2}\right] = 0, \tag{3.55}$$

we observe that

$$\mathbb{E}_\theta\left[f(X_0)\partial_x f(X_{\Delta_n}) b(X_{\Delta_n};\theta)\mathbb{E}_\theta\left(\xi_{1,2} \mid \mathcal{F}_{\Delta_n}\right)\right] = 0$$

since $\xi_{1,2} \sim \mathcal{N}(0,1/3)$ and independent of $\mathcal{F}_{\Delta_n}$ by Proposition 3.3.3.

Term 3: By applying the moment expansion (3.22) followed by Lemma 2.10.2,

$$\begin{aligned}
\mathbb{E}_\theta\left[f(X_0)\xi_{2,2}\right] &= \mathbb{E}_\theta\left[f(X_0)\mathbb{E}_\theta\left(\xi_{2,2} \mid \mathcal{F}_{\Delta_n}\right)\right] \\
&= \Delta_n \mathbb{E}_\theta\left[f(X_0)\mathcal{H}_\theta f(X_{\Delta_n})\right] + \Delta_n^{3/2} R(\Delta_n;\theta) \\
&= \Delta_n \mathbb{E}_\theta\left[f(X_0)\mathbb{E}_\theta\left(\mathcal{H}_\theta f(X_{\Delta_n}) \mid \mathcal{F}_0\right)\right] + \Delta_n^{3/2} R(\Delta_n;\theta) \\
&= \Delta_n \mu_\theta(f\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta).
\end{aligned}$$

Term 4: To expand

$$\Delta_n^{1/2}\mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta)\xi_{1,1} f(X_{\Delta_n})\right]$$

the tower property is no longer directly applicable. Instead, we again rely on the Euler-Itô expansion (3.54) of Proposition 3.3.1 and observe that

$$\Delta_n^{1/2}\mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta)\xi_{1,1} f(X_{\Delta_n})\right] =$$
$$\Delta_n \mathbb{E}_\theta\left[\left[\partial_x f(X_0) b(X_0;\theta)\right]^2 \xi_{1,1}\varepsilon_{1,1}\right] + \Delta_n^{1/2}\mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta)\xi_{1,1}\varepsilon_{2,1}\right].$$

To evaluate the first term, the isometry (3.25) implies that $\mathbb{E}_\theta\left(\xi_{1,1}\varepsilon_{1,1} \mid \mathcal{F}_0\right) = 1/2$ and, hence,

$$\begin{aligned}
\Delta_n \mathbb{E}_\theta\left[[\partial_x f(X_0) b(X_0;\theta)]^2 \xi_{1,1}\varepsilon_{1,1}\right] &= \Delta_n \mathbb{E}_\theta\left[[\partial_x f(X_0) b(X_0;\theta)]^2 \mathbb{E}_\theta\left(\xi_{1,1}\varepsilon_{1,1} \mid \mathcal{F}_0\right)\right] \\
&= \Delta_n \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right).
\end{aligned}$$

Moreover, by Hölder's inequality and the conditional moment expansion (3.17),

$$|\mathbb{E}_\theta\left(\xi_{1,1}\varepsilon_{2,1} \mid \mathcal{F}_0\right)| \le \mathbb{E}_\theta\left(\xi_{1,1}^2 \;\middle|\; \mathcal{F}_0\right)^{1/2} \mathbb{E}_\theta\left(\varepsilon_{2,1}^2 \;\middle|\; \mathcal{F}_0\right)^{1/2} = \Delta_n R(\Delta_n, X_0;\theta),$$

and, as a consequence,

$$\Delta_n^{1/2} \mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta) \mathbb{E}_\theta\left(\xi_{1,1}\varepsilon_{2,1} \mid \mathcal{F}_0\right)\right] = \Delta_n^{3/2} R(\Delta_n;\theta).$$

All in all,

$$\Delta_n^{1/2} \mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta) \xi_{1,1} f(X_{\Delta_n})\right] = \Delta_n \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right) + \Delta_n^{3/2} R(\Delta_n;\theta). \tag{3.56}$$

Term 5: By Proposition 3.3.3,

$$\mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta) \xi_{1,1} \partial_x f(X_{\Delta_n}) b(X_{\Delta_n};\theta) \mathbb{E}_\theta\left(\xi_{1,2} \mid \mathcal{F}_{\Delta_n}\right)\right] = 0.$$

Term 6: By Hölder's inequality,

$$|\mathbb{E}_\theta\left(\xi_{1,1}\xi_{2,2} \mid \mathcal{F}_0\right)| \le \mathbb{E}_\theta\left(\xi_{1,1}^2 \;\middle|\; \mathcal{F}_0\right)^{1/2} \mathbb{E}_\theta\left(\xi_{2,2}^2 \;\middle|\; \mathcal{F}_0\right)^{1/2}$$

and due to the tower property,

$$\mathbb{E}_\theta\left(\xi_{2,2}^2 \;\middle|\; \mathcal{F}_0\right) = \mathbb{E}_\theta\left[\mathbb{E}_\theta\left(\xi_{2,2}^2 \;\middle|\; \mathcal{F}_{\Delta_n}\right) \;\middle|\; \mathcal{F}_0\right] = \mathbb{E}_\theta\left[\Delta_n^2 R(\Delta_n, X_{\Delta_n};\theta) \;\middle|\; \mathcal{F}_0\right] = \Delta_n^2 R(\Delta_n, X_0;\theta)$$

and we conclude that

$$\Delta_n^{1/2} \mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta) \mathbb{E}_\theta\left(\xi_{1,1}\xi_{2,2} \mid \mathcal{F}_0\right)\right] = \Delta_n^{3/2} R(\Delta_n;\theta).$$

Term 7: For the expansion of $\mathbb{E}_\theta\left[f(X_{\Delta_n})\xi_{2,1}\right]$, we again apply the Euler-Itô expansion (3.54) of $f(X_{\Delta_n})$; firstly, by Proposition 3.3.3,

$$\begin{aligned}
\mathbb{E}_\theta\left[f(X_0)\xi_{2,1}\right] &= \mathbb{E}_\theta\left[f(X_0)\mathbb{E}_\theta\left(\xi_{2,1} \mid \mathcal{F}_0\right)\right] \\
&= \Delta_n \mu_\theta(f\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta)
\end{aligned}$$

and by Hölder's inequality, $\mathbb{E}_\theta\left(\varepsilon_{1,1}\xi_{2,1} \mid \mathcal{F}_0\right) = \Delta_n R(\Delta_n, X_0;\theta)$, from which it follows that

$$\Delta_n^{1/2} \mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta) \mathbb{E}_\theta\left(\varepsilon_{1,1}\xi_{2,1} \mid \mathcal{F}_0\right)\right] = \Delta_n^{3/2} R(\Delta_n;\theta).$$

Similarly, we see that $\mathbb{E}_\theta(\varepsilon_{2,1}\xi_{2,1}) = \Delta_n^2 R(\Delta_n;\theta)$ and, in turn,

$$\mathbb{E}_\theta\left[f(X_{\Delta_n})\xi_{2,1}\right] = \Delta_n \mu_\theta(f\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta). \tag{3.57}$$

TERM 8: By Proposition 3.3.3,

$$\mathbb{E}_\theta \left[\xi_{2,1}\partial_x f(X_{\Delta_n}) b(X_{\Delta_n};\theta)\mathbb{E}_\theta\left(\xi_{1,2} \mid \mathcal{F}_{\Delta_n}\right)\right] = 0. \tag{3.58}$$

TERM 9: Combining Hölder's inequality, the tower property and Proposition 3.3.3,

$$|\mathbb{E}_\theta\left(\xi_{2,1}\xi_{2,2}\right)| \leq \mathbb{E}_\theta\left[\mathbb{E}_\theta\left(\xi_{2,1}^2 \;\middle|\; \mathcal{F}_0\right)\right]^{1/2}\mathbb{E}_\theta\left[\mathbb{E}_\theta\left(\xi_{2,2}^2 \;\middle|\; \mathcal{F}_{\Delta_n}\right)\right]^{1/2} = \Delta_n^2 R(\Delta_n;\theta).$$

Gathering the observations of TERM 1-9, we conclude that

$$\begin{gathered}\mathbb{E}_\theta f(Y_1)f(Y_2) = \\ \mu_\theta(f^2) + \Delta_n\left(\mu_\theta(f\mathcal{L}_\theta f) + 2\mu_\theta(f\mathcal{H}_\theta f) + \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right) + \Delta_n^{3/2}R(\Delta_n;\theta).\end{gathered} \tag{3.59}$$

Step 2 To eliminate $\mathcal{H}_\theta$ from the expansion of $[\mathbb{E}_\theta f(Y_1)]^2$, $\mathbb{E}_\theta f^2(Y_1)$ and $\mathbb{E}_\theta\left[f(Y_1)f(Y_2)\right]$, respectively, we examine $\mu_\theta(\mathcal{H}_\theta f)$, $\mu_\theta(f\mathcal{H}_\theta f)$ and $\mu_\theta(\mathcal{H}_\theta f^2)$ more closely in terms of the defining property (3.24).

For the invariant distribution μ_θ and the generator of (X_t), it holds that $\mu_\theta(\mathcal{L}_\theta f) = 0$ for all $f \in \mathcal{D}_{\mathcal{A}_\theta}$; see e.g. Hansen and Scheinkman (1995). Hence,

$$\mu_\theta(\mathcal{H}_\theta f) = -\frac{1}{12}\mu_\theta\left(b^2(\,\cdot\,;\theta)\partial_x^2 f\right) \tag{3.60}$$

and it follows immediately that

$$\mu_\theta(f\mathcal{H}_\theta f) = \frac{1}{2}\mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{12}\mu_\theta\left(fb^2(\,\cdot\,;\theta)\partial_x^2 f\right). \tag{3.61}$$

Moreover, since $\partial_x f^2 = 2f\partial_x f$ and $\partial_x^2 f^2 = 2\left[f\partial_x^2 f + (\partial_x f)^2\right]$,

$$\begin{aligned}\mathcal{H}_\theta f^2(x) &= \frac{1}{2}\mathcal{L}_\theta f^2(x) - \frac{1}{12}b^2(x;\theta)\partial_x^2 f^2(x) \\ &= \frac{1}{2}a(x;\theta)\partial_x f^2(x) + \frac{1}{6}b^2(x;\theta)\partial_x^2 f^2(x) \\ &= f(x)a(x;\theta)\partial_x f(x) + \frac{1}{3}f(x)b^2(x;\theta)\partial_x^2 f(x) + \frac{1}{3}[b(x;\theta)\partial_x f(x)]^2 \\ &= f(x)\mathcal{L}_\theta f(x) - \frac{1}{6}f(x)b^2(x;\theta)\partial_x^2 f(x) + \frac{1}{3}[b(x;\theta)\partial_x f(x)]^2,\end{aligned}$$

which shows that

$$\mu_\theta(\mathcal{H}_\theta f^2) = \mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{6}\mu_\theta\left(fb^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{3}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right). \tag{3.62}$$

In turn, this enables us to write

$$
\begin{aligned}
& [\mathbb{E}_\theta f(Y_1)]^2 \\
= \;& \mu_\theta(f)^2 + \Delta_n 2\mu_\theta(f)\mu_\theta(\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta) \\
= \;& \mu_\theta(f)^2 - \Delta_n \frac{1}{6}\mu_\theta(f)\mu_\theta\left(b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \Delta_n^{3/2} R(\Delta_n;\theta),
\end{aligned}
\tag{3.63}
$$

$$
\begin{aligned}
& \mathbb{E}_\theta f^2(Y_1) \\
= \;& \mu_\theta(f^2) + \Delta_n \mu_\theta(\mathcal{H}_\theta f^2) + \Delta_n^{3/2} R(\Delta_n;\theta) \\
= \;& \mu_\theta(f^2) + \Delta_n \left(\mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{3}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right) \\
+ \;& \Delta_n^{3/2} R(\Delta_n;\theta),
\end{aligned}
\tag{3.64}
$$

and, lastly,

$$
\begin{aligned}
& \mathbb{E}_\theta f(Y_1) f(Y_2) \\
= \;& \mu_\theta(f^2) + \Delta_n \left(\mu_\theta(f\mathcal{L}_\theta f) + 2\mu_\theta(f\mathcal{H}_\theta f) + \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right) + \Delta_n^{3/2} R(\Delta_n;\theta) \\
= \;& \mu_\theta(f^2) + \Delta_n \left(2\mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right) \\
+ \;& \Delta_n^{3/2} R(\Delta_n;\theta).
\end{aligned}
\tag{3.65}
$$

Step 3 From the moment expansions (3.63)-(3.65) derived above, it follows that

$$
\begin{aligned}
\breve{a}_n(\theta)_1 &= \frac{\mathbb{E}_\theta f(Y_1) f(Y_2) - [\mathbb{E}_\theta f(Y_1)]^2}{\mathbb{V}ar_\theta f(Y_1)} \\
&= \frac{1 + \Delta_n \mathbb{V}ar_\theta f(X_0)^{-1} M_1(\theta) + \Delta_n^{3/2} R(\Delta_n;\theta)}{1 + \Delta_n \mathbb{V}ar_\theta f(X_0)^{-1} M_2(\theta) + \Delta_n^{3/2} R(\Delta_n;\theta)}
\end{aligned}
\tag{3.66}
$$

where

$$
\begin{aligned}
M_1(\theta) &= 2\mu_\theta(f\mathcal{L}_\theta f) + \frac{1}{6}\mu_\theta(f)\mu_\theta\left(b^2(\,\cdot\,;\theta)\partial_x^2 f\right) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right), \\
M_2(\theta) &= \mu_\theta(f\mathcal{L}_\theta f) + \frac{1}{6}\mu_\theta(f)\mu_\theta\left(b^2(\,\cdot\,;\theta)\partial_x^2 f\right) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{3}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right).
\end{aligned}
$$

In particular,

$$
M_1(\theta) - M_2(\theta) = \mu_\theta(f\mathcal{L}_\theta f) + \frac{1}{6}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)
$$

and by Taylor expanding the power fraction in (3.66), we obtain the expansion

$$
\begin{aligned}
\breve{a}_n(\theta)_1 &= 1 + \Delta_n \mathbb{V}ar_\theta f(X_0)^{-1}\left[M_1(\theta) - M_2(\theta)\right] + \Delta_n^{3/2} R(\Delta_n;\theta) \\
&= 1 + \Delta_n K_f(\theta) + \Delta_n^{3/2} R(\Delta_n;\theta),
\end{aligned}
\tag{3.67}
$$

where $K_f(\theta)$ is that of (3.37).

Step 4 Since $\mathbb{E}_\theta f(Y_1) = \mu_\theta(f) + \Delta_n R(\Delta_n;\theta)$, the expansion of $\breve{a}_n(\theta)_1$ in (3.67) implies that

$$\breve{a}_n(\theta)_0 = \mathbb{E}_\theta f(Y_1)\left(1 - \breve{a}_n(\theta)_1\right) = -\Delta_n K_f(\theta)\mu_\theta(f) + \Delta_n^{3/2} R(\Delta_n;\theta). \tag{3.68}$$

□

Proof of Lemma 3.5.6. In the following, we let

$$H_n(\theta) = \frac{1}{n\Delta_n}\sum_{i=2}^{n} g(\Delta_n, Y_i, Y_{i-1};\theta),$$

where $g = (g_1, g_2)^T$ is given by

$$\begin{aligned} g_1(\Delta_n, Y_i, Y_{i-1};\theta) &= f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1}), \\ g_2(\Delta_n, Y_i, Y_{i-1};\theta) &= f(Y_{i-1})\left[f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1})\right], \end{aligned} \tag{3.69}$$

corresponding to the entries of G_n.

Firstly, by applying the power expansion (3.36) of $\breve{a}_n(\theta)$ derived in Lemma 3.5.4,

$$\begin{aligned} & g_1(\Delta_n, Y_i, Y_{i-1};\theta) \\ = \; & f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1}) \\ = \; & f(Y_i) - f(Y_{i-1}) + \Delta_n K_f(\theta)\left[\mu_\theta(f) - f(Y_{i-1})\right] + \Delta_n^{3/2} R(\Delta_n, Y_{i-1};\theta) \end{aligned} \tag{3.70}$$

and, hence, by the LLN for integrated diffusions (Lemma 3.4.1),

$$\begin{aligned} & \frac{1}{n\Delta_n}\sum_{i=2}^{n} g_1(\Delta_n, Y_i, Y_{i-1};\theta) \\ = \; & \frac{1}{n\Delta_n}\left[f(Y_n) - f(Y_1)\right] + \frac{1}{n}\sum_{i=2}^{n} K_f(\theta)\left[\mu_\theta(f) - f(Y_{i-1})\right] + \Delta_n^{1/2}\frac{1}{n}\sum_{i=2}^{n} R(\Delta_n, Y_{i-1};\theta) \\ \xrightarrow{\mathbb{P}_0} \; & K_f(\theta)(\mu_\theta - \mu_0)(f). \end{aligned}$$

Conversely, for the second vector component

$$\frac{1}{n\Delta_n}\sum_{i=2}^{n} g_2(\Delta_n, Y_i, Y_{i-1};\theta), \tag{3.71}$$

the contribution from the first term is no longer asymptotically negligible and the proof more extensive. To shorten notation in the following, we let

$$\begin{aligned} \mathcal{E}_i^n &= \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta_0)\varepsilon_{1,i}, \\ \Xi_i^n &= \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta_0)\xi_{1,i}, \end{aligned}$$

which, in turn, enables us to write the expansions of Section 3.3 under the true probability measure $\mathbb{P}_0$ as

$$f(X_{t_i^n}) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\mathcal{E}_i^n + \varepsilon_{2,i}, \tag{3.72}$$

$$f(Y_i) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\Xi_i^n + \xi_{2,i}. \tag{3.73}$$

By inserting the expansion (3.73) directly into (3.69) and applying the power expansion of $\breve{a}_n(\theta)$, we find that

$$g_1(\Delta_n, Y_i, Y_{i-1}; \theta) = f(X_{t_{i-1}^n}) - f(X_{t_{i-2}^n}) +$$
$$\Delta_n K_f(\theta)\left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] + \Delta_n^{1/2}\left(\Xi_i^n - \Xi_{i-1}^n\right) + \mathcal{R}_1(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0; \theta), \tag{3.74}$$

where the remainder term $\mathcal{R}_1$ has a semi-explicit representation of the form

$$\mathcal{R}_1(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0; \theta) =$$
$$(\xi_{2,i} - \xi_{2,i-1}) - \Delta_n^{3/2} K_f(\theta)\Xi_{i-1}^n - \Delta_n K_f(\theta)\xi_{2,i-1} + \Delta_n^{3/2} R(\Delta_n, Y_{i-1}; \theta). \tag{3.75}$$

In turn, this enables us to expand

$$g_2(\Delta_n, Y_i, Y_{i-1}; \theta) = \left(f(X_{t_{i-2}^n}) + \Delta_n^{1/2}\Xi_{i-1}^n + \xi_{2,i-1}\right) g_1(\Delta_n, Y_i, Y_{i-1}; \theta)$$

by decomposing the functional increment of (X_t) in (3.74) as

$$f(X_{t_{i-1}^n}) - f(X_{t_{i-2}^n}) = \Delta_n^{1/2}\mathcal{E}_{i-1}^n + \varepsilon_{2,i-1} = \Delta_n \mathcal{L}_0 f(X_{t_{i-2}^n}) + A_{i-1}(\theta_0) + M_{i-1}(\theta_0),$$

where the first expression follows from (3.72) and the latter holds by Itô's formula with

$$A_i(\theta) = \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\mathcal{L}_\theta f(X_s) - \mathcal{L}_\theta f(X_{t_{i-1}^n})\right] \mathrm{d}s,$$
$$M_i(\theta) = \int_{(i-1)\Delta_n}^{i\Delta_n} \partial_x f(X_s) b(X_s;\theta) dB_s.$$

Specifically, we obtain the Δ-expansion

$$g_2(\Delta_n, Y_i, Y_{i-1}; \theta) = \sum_{k=1}^{3} g_2^{(k)}(\Delta_n, Y_i, Y_{i-1}; \theta), \tag{3.76}$$

where each constituent $g_2^{(k)}$, $k = 1, 2, 3$, has a representation

$$\begin{aligned}
& g_2^{(1)}(\Delta_n, Y_i, Y_{i-1}; \theta) \\
= \; & f(X_{t_{i-2}^n}) \cdot g_1(\Delta_n, Y_i, Y_{i-1}; \theta) \\
= \; & \Delta_n f(X_{t_{i-2}^n})\mathcal{L}_0 f(X_{t_{i-2}^n}) + f(X_{t_{i-2}^n}) M_{i-1}(\theta_0) + \Delta_n K_f(\theta) f(X_{t_{i-2}^n})\left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] \\
+ \; & \mathcal{R}_2^{(1)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0; \theta)
\end{aligned}$$

for a remainder term

$$\mathcal{R}_2^{(1)}(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=$$
$$f(X_{t_{i-2}^n})A_{i-1}(\theta_0)+\Delta_n^{1/2}f(X_{t_{i-2}^n})\left(\Xi_i^n-\Xi_{i-1}^n\right)+f(X_{t_{i-2}^n})\cdot\mathcal{R}_1(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta),$$

$$\begin{aligned} g_2^{(2)}(\Delta_n,Y_i,Y_{i-1};\theta) &= \Delta_n^{1/2}\cdot\Xi_{i-1}^n\cdot g_1(\Delta_n,Y_i,Y_{i-1};\theta) \\ &= \Delta_n\left(\mathcal{E}_{i-1}^n-\Xi_{i-1}^n\right)\Xi_{i-1}^n+\mathcal{R}_2^{(2)}(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta) \end{aligned}$$

where

$$\mathcal{R}_2^{(2)}(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=\Delta_n^{1/2}\Xi_{i-1}^n\varepsilon_{2,i-1}+$$
$$\Delta_n^{3/2}\Xi_{i-1}^nK_f(\theta)\left[\mu_\theta(f)-f(X_{t_{i-2}^n})\right]+\Delta_n\Xi_i^n\Xi_{i-1}^n+\Delta_n^{1/2}\Xi_{i-1}^n\mathcal{R}_1(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta),$$

and

$$g_2^{(3)}(\Delta_n,Y_i,Y_{i-1};\theta)=\xi_{2,i-1}\cdot g_1(\Delta_n,Y_i,Y_{i-1};\theta)=\mathcal{R}_2^{(3)}(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)$$

respectively.

Collecting the terms,

$$\begin{aligned} & g_2(\Delta_n,Y_i,Y_{i-1};\theta) \\ = & \sum_{k=1}^3 g_2^{(k)}(\Delta_n,Y_i,Y_{i-1};\theta) \\ = & \Delta_nf(X_{t_{i-2}^n})\mathcal{L}_0f(X_{t_{i-2}^n})+f(X_{t_{i-2}^n})M_{i-1}(\theta_0)+\Delta_nK_f(\theta)f(X_{t_{i-2}^n})\left[\mu_\theta(f)-f(X_{t_{i-2}^n})\right] \\ + & \Delta_n\left(\mathcal{E}_{i-1}^n-\Xi_{i-1}^n\right)\Xi_{i-1}^n+\mathcal{R}_2(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta), \end{aligned} \tag{3.77}$$

where the (somewhat dramatic) remainder term

$$\mathcal{R}_2(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=\sum_{k=1}^3\mathcal{R}_2^{(k)}(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta).$$

At this point, tedious reasoning based on Lemma 9 in Genon-Catalot and Jacod (1993) ensures that

$$\frac{1}{n\Delta_n}\sum_{i=2}^n\mathcal{R}_2(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=o_{\mathbb{P}_0}(1) \tag{3.78}$$

and, furthermore,

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n\mathcal{R}_2(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=o_{\mathbb{P}_0}(1) \tag{3.79}$$

under the additional rate assumption $n\Delta_n^2 \to 0$.[3] To see that the strong rate assumption $n\Delta_n^2 \to 0$ is necessary to obtain (3.79) we can, e.g., consider the last remainder term in (3.75) since

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^n \Delta_n^{3/2} R(\Delta_n, Y_{i-1}; \theta) = \sqrt{n\Delta_n^2} \cdot \frac{1}{n} \sum_{i=2}^n R(\Delta_n, Y_{i-1}; \theta).$$

As the proofs of (3.78) and (3.79) are both very long and not very insightful, they are omitted.

To determine the limit in probability of (3.71), we consider each term in the Δ-expansion (3.77) separately: by Lemma 2.3.1,

$$\frac{1}{n} \sum_{i=2}^n f(X_{t_{i-2}^n}) \mathcal{L}_0 f(X_{t_{i-2}^n}) \xrightarrow{\mathbb{P}_0} \mu_0(f \mathcal{L}_0 f)$$

and, analogously, we have

$$\frac{1}{n} \sum_{i=2}^n K_f(\theta) f(X_{t_{i-2}^n}) \left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] \xrightarrow{\mathbb{P}_0} K_f(\theta) \left[\mu_0(f)\mu_\theta(f) - \mu_0(f^2)\right].$$

Furthermore, by construction

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left((\mathcal{E}_i^n - \Xi_i^n) \Xi_i^n \mid \mathcal{F}_{i-1}^n \right) = \frac{1}{n} \sum_{i=1}^n [\partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n}; \theta_0)]^2 \mathbb{E}_0 \left((\varepsilon_{1,i} - \xi_{1,i}) \xi_{1,i} \mid \mathcal{F}_{i-1}^n \right)$$

and since $\xi_{1,i} \sim \mathcal{N}(0, 1/3)$, the isometry (3.25) implies that

$$\mathbb{E}_0 \left((\varepsilon_{1,i} - \xi_{1,i}) \xi_{1,i} \mid \mathcal{F}_{i-1}^n \right) = \mathbb{E}_0 \left((\varepsilon_{1,i} - \xi_{1,i}) \xi_{1,i} \right) = \frac{1}{6} \tag{3.80}$$

and

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E}_0 \left((\mathcal{E}_i^n - \Xi_i^n) \Xi_i^n \mid \mathcal{F}_{i-1}^n \right) \xrightarrow{\mathbb{P}_0} \frac{1}{6} \mu_0 \left([b(\,\cdot\,; \theta_0) \partial_x f]^2 \right).$$

Therefore, by observing from the construction of $\varepsilon_{1,i}$ in (3.42) and $\xi_{1,i}$ in (3.47) that the difference $\varepsilon_{1,i} - \xi_{1,i}$ remains Gaussian, Hölder's inequality implies that

$$\begin{aligned} & \frac{1}{n^2} \sum_{i=1}^n \mathbb{E}_0 \left((\mathcal{E}_i^n - \Xi_i^n)^2 (\Xi_i^n)^2 \mid \mathcal{F}_{i-1}^n \right) \\ = & \frac{1}{n^2} \sum_{i=1}^n [\partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n}; \theta_0)]^4 \cdot \mathbb{E}_0 \left((\varepsilon_{1,i} - \xi_{1,i})^2 \xi_{1,i}^2 \,\middle|\, \mathcal{F}_{i-1}^n \right) \\ = & o_{\mathbb{P}_0}(1), \end{aligned}$$

[3] The latter observation is redundant for the proof at hand, however necessary to obtain asymptotic normality in Theorem 3.5.7. We state it here for convenience.

and

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{E}_i^n - \Xi_i^n\right)\Xi_i^n \xrightarrow{\mathbb{P}_0} \frac{1}{6}\mu_0\left([b(\,\cdot\,;\theta_0)\partial_x f]^2\right)$$

by Lemma 9 in Genon-Catalot and Jacod (1993).

By the same argument,

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n} f(X_{t_{i-1}^n})M_i(\theta_0) = o_{\mathbb{P}_0}(1),$$

where we apply that $\mathbb{E}_0\left(M_i(\theta_0)\;\middle|\;\mathcal{F}_{i-1}^n\right) = 0$ and, with $h(x) = \partial_x f(x)b(x;\theta_0)$, it follows from the conditional Itô isometry, Tonelli's theorem and Lemma 2.10.2 that

$$\begin{aligned}
\mathbb{E}_0\left(M_i^2(\theta_0)\;\middle|\;\mathcal{F}_{i-1}^n\right) &= \mathbb{E}_0\left(\int_{(i-1)\Delta_n}^{i\Delta_n} h^2(X_s)\,\mathrm{d}s\;\middle|\;\mathcal{F}_{i-1}^n\right)\\
&= \int_{(i-1)\Delta_n}^{i\Delta_n} \mathbb{E}_0\left(h^2(X_s)\;\middle|\;\mathcal{F}_{i-1}^n\right)\mathrm{d}s\\
&= \int_0^{\Delta_n}\left[h^2(X_{t_{i-1}^n}) + u\cdot R(u, X_{t_{i-1}^n};\theta_0)\right]\mathrm{d}u\\
&= \Delta_n h^2(X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta_0)
\end{aligned}$$

and, therefore,

$$\begin{aligned}
&\frac{1}{n^2\Delta_n^2}\sum_{i=1}^{n}\mathbb{E}_0\left(f^2(X_{t_{i-1}^n})M_i^2(\theta_0)\;\middle|\;\mathcal{F}_{i-1}^n\right)\\
=\;& \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^{n} f^2(X_{t_{i-1}^n})h^2(X_{t_{i-1}^n}) + \frac{1}{n^2}\sum_{i=1}^{n} R(\Delta_n, X_{t_{i-1}^n};\theta_0)\\
=\;& o_{\mathbb{P}_0}(1).
\end{aligned}$$

Gathering our observations,

$$\begin{gathered}
\frac{1}{n\Delta_n}\sum_{i=2}^{n} g_2(\Delta_n, Y_i, Y_{i-1};\theta) \xrightarrow{\mathbb{P}_0}\\
\mu_0(f\mathcal{L}_0 f) + \frac{1}{6}\mu_0\left([b(\,\cdot\,;\theta_0)\partial_x f]^2\right) - K_f(\theta)\left[\mu_0(f^2) - \mu_0(f)\mu_\theta(f)\right],
\end{gathered}$$

which verifies (3.39).

To identify the limit of $\partial_{\theta^T} H_n(\theta)$, we write

$$H_n(\theta) = \frac{1}{n\Delta_n}\sum_{i=2}^{n} Z_{i-1}\left[f(Y_i) - Z_{i-1}^T \breve{a}_n(\theta)\right],$$

which implies that

$$\partial_{\theta^T} H_n(\theta) = -\frac{1}{n\Delta_n}\sum_{i=2}^{n} Z_{i-1}Z_{i-1}^T \partial_{\theta^T}\breve{a}_n(\theta) = Z_n(f)A_n(\theta),$$

where $Z_n(f) := \frac{1}{n}\sum_{i=2}^n Z_{i-1}Z_{i-1}^T$ and $A_n(\theta) := -\Delta_n^{-1}\partial_{\theta^T}\breve{a}_n(\theta)$. By Lemma 3.4.1,

$$Z_n(f) \xrightarrow{\mathbb{P}_0} Z(f) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix}$$

and applying the power expansion (3.36) of $\breve{a}_n(\theta)$,

$$A_n(\theta) = \partial_{\theta^T}\begin{pmatrix} K_f(\theta)\mu_\theta(f) \\ -K_f(\theta) \end{pmatrix} + \Delta_n^{1/2}\partial_{\theta^T}R(\Delta_n;\theta) \to \partial_{\theta^T}\begin{pmatrix} K_f(\theta)\mu_\theta(f) \\ -K_f(\theta) \end{pmatrix} =: A(\theta).$$

Hence, it follows that

$$\partial_{\theta^T} H_n(\theta) \xrightarrow{\mathbb{P}_0} Z(f)A(\theta) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix}\begin{pmatrix} \partial_{\theta_1}\left[K_f(\theta)\mu_\theta(f)\right] & \partial_{\theta_2}\left[K_f(\theta)\mu_\theta(f)\right] \\ -\partial_{\theta_1}K_f(\theta) & -\partial_{\theta_2}K_f(\theta) \end{pmatrix}$$

by the continuous mapping theorem. To argue that the convergence is uniform over $\mathcal{M}$ under Condition 3.5.5, note that

$$\begin{aligned} \|\partial_{\theta^T}H_n(\theta) - Z(f)A(\theta)\| &= \|Z_n(f)A_n(\theta) - Z(f)A(\theta)\| \\ &= \|Z_n(f)A_n(\theta) - Z_n(f)A(\theta) + Z_n(f)A(\theta) - Z(f)A(\theta)\| \\ &\leq \|Z_n(f)[A_n(\theta) - A(\theta)]\| + \|[Z_n(f) - Z(f)]A(\theta)\| \end{aligned}$$

and, in particular,

$$\sup_{\theta\in\mathcal{M}}\|\partial_{\theta^T}H_n(\theta) - Z(f)A(\theta)\| \leq \|Z_n(f)\|\sup_{\theta\in\mathcal{M}}\|A_n(\theta) - A(\theta)\| + \|Z_n(f) - Z(f)\|\sup_{\theta\in\mathcal{M}}\|A(\theta)\|.$$

Therefore, (3.40) follows by observing that

$$\sup_{\theta\in\mathcal{M}}\|A_n(\theta) - A(\theta)\| = \Delta_n^{1/2}\sup_{\theta\in\mathcal{M}}\|\partial_{\theta^T}R(\Delta_n;\theta)\| \leq_{C(\mathcal{M})} \Delta_n^{1/2} \to 0$$

and applying continuity of $\|\cdot\|$ and $\theta \mapsto A(\theta)$, respectively. □

Proof of Theorem 3.5.7. We continue with the notation from the proof of Lemma 3.5.6. Existence of a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ follows directly from Lemma 3.5.6 and Theorem 1.58 in M. Sørensen (2012). Moreover, asymptotic normality of $\hat{\theta}_n$ can be shown by Taylor expanding $H_n(\theta)$ once we establish that

$$\sqrt{n\Delta_n}\cdot H_n(\theta_0) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2(0, V_0(f)). \tag{3.81}$$

From the Δ-expansion of $g_1(\Delta_n, Y_i, Y_{i-1}; \theta)$ in (3.70), it follows that

$$
\begin{aligned}
& \frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^{n} g_1(\Delta_n, Y_i, Y_{i-1}; \theta_0) \\
= \;& \frac{1}{\sqrt{n\Delta_n}} [f(Y_n) - f(Y_1)] + \sqrt{n\Delta_n} \left(\frac{1}{n} \sum_{i=2}^{n} f_1^*(Y_{i-1}) \right) + \sqrt{n\Delta_n^2} \cdot \frac{1}{n} \sum_{i=2}^{n} R(\Delta_n, Y_{i-1}; \theta_0) \\
= \;& \sqrt{n\Delta_n} \left(\frac{1}{n} \sum_{i=2}^{n} f_1^*(Y_{i-1}) \right) + o_{\mathbb{P}_0}(1) \\
\xrightarrow{\mathscr{D}_0} \;& \mathcal{N}\left(0, \mu_0\left([\partial_x U_0(f_1^*) b(\,\cdot\,; \theta_0)]^2\right)\right),
\end{aligned}
$$

where convergence in law is true by Lemma 3.4.2 since $f_1^* \in \mathscr{H}_0^2$.

In turn, our proof that

$$
\frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^{n} g_2(\Delta_n, Y_i, Y_{i-1}; \theta_0) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mu_0\left(\left[\partial_x U_0(f_2^*) + f \partial_x f\right]^2 b^2(\,\cdot\,; \theta_0)\right)\right) \tag{3.82}
$$

is based on the observation that

$$
\frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^{n} \Delta_n \left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right) \Xi_{i-1}^n = \frac{1}{6} \frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^{n} \Delta_n [\partial_x f(X_{t_{i-2}^n}) b(X_{t_{i-2}^n}; \theta_0)]^2 + o_{\mathbb{P}_0}(1), \tag{3.83}
$$

which follows from Lemma 9 in Genon-Catalot and Jacod (1993) by applying that

$$
\left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right) \Xi_{i-1}^n = [\partial_x f(X_{t_{i-2}^n}) b(X_{t_{i-2}^n}; \theta_0)]^2 \left(\varepsilon_{1,i-1} - \xi_{1,i-1}\right) \xi_{1,i-1},
$$

and, in particular,

$$
\mathbb{E}_0\left(\left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right) \Xi_{i-1}^n \;\middle|\; \mathcal{F}_{i-2}^n\right) = \frac{1}{6} [\partial_x f(X_{t_{i-2}^n}) b(X_{t_{i-2}^n}; \theta_0)]^2.
$$

Combining the equivalence (3.83) with the Δ-expansion of $g_2(\Delta_n, Y_i, Y_{i-1}; \theta_0)$ in (3.77) and the vanishing condition (3.79) of the corresponding remainder term, we see that

$$
\begin{aligned}
& \frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^{n} g_2(\Delta_n, Y_i, Y_{i-1}; \theta_0) \\
= \;& \sqrt{n\Delta_n} \left(\frac{1}{n} \sum_{i=2}^{n} f_2^*(X_{t_{i-2}^n}) \right) + \frac{1}{\sqrt{n\Delta_n}} \sum_{i=2}^{n} f(X_{t_{i-2}^n}) M_{i-1}(\theta_0) + o_{\mathbb{P}_0}(1) \\
= \;& \sqrt{n\Delta_n} \left(\frac{1}{n} \sum_{i=1}^{n} f_2^*(X_{t_{i-1}^n}) \right) + \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^{n} f(X_{t_{i-1}^n}) M_i(\theta_0) + o_{\mathbb{P}_0}(1),
\end{aligned} \tag{3.84}
$$

and to gather the non-negligible terms in (3.84), we initially observe that

$$\begin{aligned}
&\frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n} f_2^*(X_s)\,\mathrm{d}s \\
=\;& \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n\int_{(i-1)\Delta_n}^{i\Delta_n} f_2^*(X_s)\,\mathrm{d}s \\
=\;& \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right)+\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n\int_{(i-1)\Delta_n}^{i\Delta_n}\left[f_2^*(X_s)-f_2^*(X_{t_{i-1}^n})\right]\,\mathrm{d}s \quad (3.85)\\
=\;& \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right)+o_{\mathbb{P}_0}(1),
\end{aligned}$$

where we only use that $f_2^*\in\mathcal{C}_p^2(S)$.[4] Furthermore, since $f_2^*\in\mathscr{H}_0^2$ under Condition 3.5.5, Proposition 2.3.3 implies that

$$\mathcal{L}_0\left(U_0(f_2^*)\right)=-f_2^* \tag{3.86}$$

and, therefore, by Itô's lemma

$$\begin{aligned}
U_0(f_2^*)(X_t) &= U_0(f_2^*)(X_0)+\int_0^t\mathcal{L}_0(U_0(f_2^*))(X_s)\,\mathrm{d}s+\int_0^t\partial_xU_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s\\
&= U_0(f_2^*)(X_0)-\int_0^t f_2^*(X_s)\,\mathrm{d}s+\int_0^t\partial_xU_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s.
\end{aligned}$$

As a consequence,

$$\begin{aligned}
\sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right) &= \frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n} f_2^*(X_s)\,\mathrm{d}s+o_{\mathbb{P}_0}(1)\\
&= \frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n}\partial_xU_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s+o_{\mathbb{P}_0}(1)\\
&= \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n\int_{(i-1)\Delta_n}^{i\Delta_n}\partial_xU_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s+o_{\mathbb{P}_0}(1),
\end{aligned}$$

$$\begin{aligned}
&\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n g_2(\Delta_n,Y_i,Y_{i-1};\theta_0)\\
=\;& \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right)+\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n f(X_{t_{i-1}^n})M_i(\theta_0)+o_{\mathbb{P}_0}(1)\\
=\;& \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n\int_{(i-1)\Delta_n}^{i\Delta_n}\left[\partial_xU_0(f_2^*)(X_s)+f(X_{t_{i-1}^n})\partial_xf(X_s)\right]b(X_s;\theta_0)dB_s+o_{\mathbb{P}_0}(1),
\end{aligned}$$

[4] A proof that the latter term in (3.85) is, in fact, asymptotically negligible under $\mathbb{P}_0$ is contained in the proof of Proposition 2.3.4.

and, at this point, the asymptotic normality in (3.82) can be shown by applying the central limit theorem for martingale difference arrays in Häusler and Luschgy (2015).[5]

The proof of joint normality can be carried out as in Chapter 2 by applying the Cramér-Wold device. □

3.8 Appendix B: Auxiliary results

Lemma 3.8.1. *Let $(X_t)_{t\geq 0}$ be a continuous semimartingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ and suppose that $(H_t)_{t\geq 0}$ is $(\mathcal{F}_t)$-adapted and continuous. For any $t \geq t^* \geq 0$,*

$$\int_{t^*}^{t} \left(\int_{t^*}^{s} H_u dX_u \right) \mathrm{d}s = \int_{t^*}^{t} (t-s) H_s dX_s.$$

Proof. Let $Z_t = \int_0^t H_s dX_s$. By stochastic integration-by-parts,

$$tZ_t = \int_0^t s dZ_s + \int_0^t Z_s \,\mathrm{d}s$$

and, in particular,

$$\int_0^t Z_s \,\mathrm{d}s = \int_0^t \left(\int_0^s H_u dX_u \right) \mathrm{d}s = \int_0^t (t-s) H_s dX_s,$$

which verifies the result for $t^* = 0$. For $t^* > 0$,

$$\begin{aligned}
\int_{t^*}^{t} \left(\int_{t^*}^{s} H_u dX_u \right) \mathrm{d}s &= \int_{t^*}^{t} (Z_s - Z_{t^*})\,\mathrm{d}s \\
&= \int_0^t (Z_s - Z_{t^*})\,\mathrm{d}s - \int_0^{t^*} (Z_s - Z_{t^*})\,\mathrm{d}s \\
&= \int_0^t (t-s) H_s dX_s - \int_0^{t^*} (t^*-s) H_s dX_s - Z_{t^*}(t-t^*), \qquad (3.87)
\end{aligned}$$

and the result follows by decomposing the middle term

$$\begin{aligned}
\int_0^{t^*} (t^*-s) H_s dX_s &= \int_0^{t^*} [t - s - (t-t^*)] H_s dX_s \\
&= \int_0^{t^*} (t-s) H_s dX_s - Z_{t^*}(t-t^*)
\end{aligned}$$

and re-inserting into (3.87). □

[5]These exact derivations can be found on pp. 41-43.

4 One-Factor Models for Diversified Stock Indices with High-Frequency Observations

Emil S. Jørgensen
University of Copenhagen

Abstract. In this paper we construct a class of continuous-time stochastic volatility models aimed at modeling the dynamics of diversified stock indices. The models are of parametric diffusion-type and are driven by a single Brownian motion that models the non-diversifiable risk of the underlying market. For the construction we utilize the concept of stochastic market time and, in particular, the base process and the random time change are dependent processes in our setup. Our emphasis is on high-frequency econometric issues related to the model class. We propose a two-step method for estimating the finite-dimensional parameter and construct a simulation-based test for the implicit one-factor hypothesis for a large class of continuous Itô semimartingales with stochastic volatility. The one-factor test is based on a nonparametric measure of instantaneous leverage effect, where the one-factor model corresponds to perfect negative correlation. We illustrate the methodology using simulated data, as well as high-frequency observations of the S&P 500.

Keywords: *Diffusion process, diversification, high-frequency data, leverage effect, one-factor model, stochastic volatility, time change.*

4.1 Introduction

A central concept in modern portfolio theory is the so-called *numéraire* or *growth optimal* portfolio which, for a given financial market, is characterized as the portfolio that attains the maximum long-term growth rate; see e.g. Platen (2011). For investment purposes, this particular characterization shows that any investor with the objective of maximizing her or his long-term wealth should invest according to the optimal weighting described by the numéraire. Unfortunately, typical semimartingale restrictions on the underlying market dynamics lead to closed-form expressions for the optimal weights that are difficult to estimate from available time series; see e.g. Platen (2002) for explicit derivations in the case of diffusion models and Fan, Y. Li, and Yu (2012) for related results on estimation of the instantaneous volatility matrix. The practical difficulty concerning estimation of the optimal weighting has motivated a stream of literature devoted to approximating the numéraire portfolio by means of a different methodology.[1]

Practitioners of long-term, risk aversive investment strategies typically aim to construct a large portfolio in such a way that

(i) the individual risks that drive the underlying assets are largely uncorrelated,

(ii) each asset is given a small weight fraction to ensure robustness against market downturns,

and recent developments in quantitative finance have shown that naive *diversification* in the sense of (ii) does, in fact, provide a reasonable approximation of the numéraire portfolio *if* the market is well-securitized in a formalized variant of (i). A general diversification theorem was established in Platen and Heath (2006) for a large class of semimartingale-based portfolios and Platen and Rendek (2012a) contains a similar result for the special case of equi-weighted indices of financial markets.

A liquid class of tradeable portfolios is provided by market capitalization weighted indices (MCIs) such as the Danish OMXC20CAP or the S&P 500. Therefore, for long-term investors, e.g. asset managers or pension funds, it is of great interest to know whether or not the market capitalization weighting provides a reasonable approximation to the corresponding numéraire portfolio. By the diversification theorem (Theorem 10.6.5) in Platen and Heath (2006), the answer to this question is affirmative as long as the index is composed of a large number of constituents and the underlying market well-securitized. Whereas the small number of stocks that constitute the OMXC20CAP seem to pose a potential diversification problem, we expect the S&P 500 to satisfy both (i) and (ii) to a large extent. The market condition (i) has spiked considerable interest into the construction of diversified indices with highly securitized markets and, e.g., Le and Platen (2006) construct a so-called world stock index based on diversification of 104 existing sector indices and argue that it leads to a reasonable approximation of the numéraire. The fact that diversification provides a *feasible* approach

[1] Besides its apparent role in portfolio optimization, the numéraire portfolio also serves as benchmark for a general approach to derivative pricing in Platen (2006) and Platen and Heath (2006).

to approximating the numéraire portfolio motivates a general study of diversified portfolio dynamics and, in particular, the dynamics of diversified stock indices which we examine in this paper.

Building on earlier work by Ignatieva and Platen (2012) and Platen and Rendek (2012b), this paper considers a class of time-changed square-root diffusion processes for modeling the normalized index dynamics. A square-root diffusion coefficient was identified by Ignatieva and Platen (2012) as a good fit for the stock index constructed in Le and Platen (2006), and the extension by stochastic time change was proposed by Platen and Rendek (2012b) to provide a better match for the stylized facts. The model is driven by a single Brownian motion that represents the non-diversifiable market risk and falls within the class of one-factor continuous-time stochastic volatility models.

Our emphasis will be on high-frequency econometric issues related to the model class. In particular, we consider estimation of the unknown model parameters and construct a simulation-based test for the implicit one-factor hypothesis for a large class of continuous Itô semimartingales with stochastic volatility. The latter is based on a normalized measure of instantaneous leverage effect, where the one-factor model corresponds to perfect negative correlation.

The structure of the remainder of the paper is as follows. In Section 4.2 we define a class of time-changed square-root diffusion processes. The time-changed model has a useful characterization as the observable component of a one-factor bivariate diffusion model which we emphasize in Section 4.2.2. Section 4.3 is devoted to a preliminary study of the model-implied volatility dynamics and leverage effects. In Section 4.4 we consider estimation of the unknown model parameters specifically aimed at the availability of high-frequency data. Section 4.5 deals with the construction of a simulation-based one-factor test for a large class of continuous-time stochastic volatility models. Important findings from an extensive Monte Carlo study are summarized in Section 4.6 and an empirical study of pre-processed 5-minute data of the S&P 500 is contained in Section 4.7. Section 4.8 concludes.

4.2 One-factor index models

In this section we construct a class of continuous-time stochastic volatility models for modeling the dynamics of diversified stock indices. Due to diversification, these indices often display an average exponential growth which we model as a separate deterministic process. Section 4.2.1 presents a model (S_t) for the normalized index dynamics based on the mathematical concept of *time change* and in Section 4.2.2 we characterize (S_t) as the observable component of a partially observed bivariate diffusion model.

4.2.1 Time-change construction

The use of time change as a means to construct stochastic volatility models in finance is widely accepted; see e.g. A. Veraart and Winkel (2010) for a concise review. For any continuous-time process (X_t) and an increasing, nonnegative process (T_t) defined on a probability space

$(\Omega, \mathcal{F}, \mathbb{P})$, the process

$$S_t(\omega) = X_{T_t(\omega)}(\omega) \tag{4.1}$$

is said to be a *time-changed process* with *base process* (X_t) and *time change* (T_t).

To ensure that the time-changed process (S_t) remains tractable for a given (X_t), additional restrictions on (T_t) are typically imposed. The following condition ensures that we can apply standard results from stochastic calculus throughout the paper.

Definition 4.2.1. A stochastic process $T = (T_t)$ is said to be a *time change* w.r.t. a given filtration $\mathscr{G} = (\mathcal{G}_t)$ if each of the following properties hold:

- $T_0 = 0$,
- $t \mapsto T_t(\omega)$ is continuous and increasing,
- for every t, the random variable T_t constitutes a $\mathscr{G}$-stopping time.

In particular, the stopping-time assumption in Definition 4.2.1 allows us to construct a new filtration $\mathscr{G}_T = (\mathcal{G}_{T_t})$ composed of the σ-fields of observable events up to time T_t; see e.g. Chapter 7, Kallenberg (2002).

The idea of replacing t-time by a stochastic market time goes back to Clark (1973), who was the first to model asset price dynamics using time-changed Brownian motion. His approach was later extended by Ané and Geman (2000) to include a more general definition of market time. In this paper, we model the time evolution of diversified stock indices as an integral w.r.t. a latent market activity process (M_t) having values in $\mathbb{R}_+ := (0, \infty)$, i.e. we define

$$\tau_t(\omega) = \int_0^t M_s(\omega)\,\mathrm{d}s \tag{4.2}$$

and we suppose throughout that (M_t) is a time-homogeneous diffusion process that satisfies the following regularity conditions:

Condition 4.2.2. *The stochastic differential equation*

$$dM_t = a(M_t)dt + b(M_t)dB_t, \;\; M_0 \sim \mu \tag{4.3}$$

where (B_t) denotes a $(\mathbb{P}, \mathscr{F})$-Brownian motion w.r.t. some filtration $\mathscr{F} = (\mathcal{F}_t)$, has a unique strong solution (M_t) *such that*

- (M_t) *is strictly stationary under* $\mathbb{P}$,
- $\int_0^\infty M_s\,\mathrm{d}s = \infty$, $\mathbb{P}$-*almost surely.*

The integral assumption $\int_0^\infty M_s \, \mathrm{d}s = \infty$ provides a natural growth condition on the market time τ and note that under Condition 4.2.2, τ is continuous, increasing and $\mathscr{F}$-adapted with $\tau_0 = 0$. The use of time-changes obtained by integrating over a positive stochastic process has been studied by e.g. Carr et al. (2003) for the construction of Lévy processes with stochastic volatility.

To construct the base process (X_t), we let

$$\rho_t = \inf\{s : \tau_s > t\} = \{s : \tau_s = t\}, \tag{4.4}$$

corresponding to *inverse market time*. By doing so,

$$\tau_{\rho_t} = \rho_{\tau_t} = t$$

and clearly the stochastic process $\rho = (\rho_t)$ is continuous, increasing, and constitutes a family of $\mathscr{F}$-stopping times. The latter follows by observing that

$$\{\rho_t \leq u\} = \{t \leq \tau_u\} \in \mathcal{F}_u$$

for any t and all $u \geq 0$ and, hence, we see that ρ defines a time change w.r.t. the market filtration $\mathscr{F}$ in accordance with Definition 4.2.1.

We model $X_t = S_{\rho_t}$ as a strong solution of the stochastic differential equation

$$dX_t = \beta(1 - X_t)dt + \sqrt{X_t} dW_t, \quad X_0 = x_0 \tag{4.5}$$

where the underlying filtered space is $(\Omega, \mathcal{F}, \mathscr{F}_\rho, \mathbb{P})$ and (W_t) denotes the following carefully constructed $(\mathbb{P}, \mathscr{F}_\rho)$-Brownian motion: since (B_t) and (M_t) are both adapted to $\mathscr{F}$, the integral process

$$I_t = \int_0^t \sqrt{M_s} dB_s$$

defines a continuous $(\mathbb{P}, \mathscr{F})$-local martingale null at zero. Therefore, by applying the Dubins-Schwarz representation of continuous local martingales (Theorem V.1.6, Revuz and Yor (1999)), the time-changed integral

$$I_{\rho_t} = \int_0^{\rho_t} \sqrt{M_s} dB_s =: W_t \tag{4.6}$$

defines a Brownian motion w.r.t. the inverse market filtration $\mathscr{F}_\rho$.

At this point, we obtain our class of one-factor stochastic volatility models

$$S_t = X_{\tau_t} \tag{4.7}$$

by time-changing (X_t) with the $\mathscr{F}_\rho$-time change τ, corresponding to market time. To see that τ constitutes a time change w.r.t. $\mathscr{F}_\rho$, it suffices to observe that

$$\{\tau_t \leq u\} = \{t \leq \rho_u\} \in \mathcal{F}_{\rho_u}$$

for any t and all $u \geq 0$. Note that, by construction, the time-changed process (S_t) in (4.7) will be adapted to the original market filtration $\mathscr{F}$ and, in particular, we have the explicit relation

$$W_{\tau_t} = I_t = \int_0^t \sqrt{M_s} dB_s, \tag{4.8}$$

which is true up to indistinguishability under $\mathbb{P}$. The relation (4.8) captures the basic assumption that diversified stock indices are driven only by a single Brownian motion that models non-diversifiable market risk.

Remark 4.2.3. Many continuous-time models with a natural time-change representation are defined in such a way that the base process and the time change are *independent*; see e.g. Clark (1973), Madan and Seneta (1990) and Barndorff-Nielsen (1997) for examples of Lévy processes, or Ané and Geman (2000) and Carr et al. (2003) for models with stochastic volatility. The Dubins-Schwarz construction of (W_t) in (4.6) implies that the base process (X_t) and the stochastic time change τ are dependent processes in our setup.

4.2.2 Stochastic differential form

One way to view (S_t) is as the observable marginal of a bivariate diffusion model for (S, M). In the following, we rely on auxiliary results in Revuz and Yor (1999) for time-changed Lebesgue-Stieltjes integrals (Proposition V.1.4) and time-changed Itô integrals (Proposition V.1.5).

Up to $\mathbb{P}$-indistinguishability, the time-changed process

$$\begin{aligned} X_{\tau_t} &= x_0 + \beta \int_0^{\tau_t} (1 - X_s) \, \mathrm{d}s + \int_0^{\tau_t} \sqrt{X_s} dW_s \\ &= x_0 + \beta \int_0^t (1 - X_{\tau_s}) \, \mathrm{d}\tau_s + \int_0^t \sqrt{X_{\tau_s}} dW_{\tau_s}, \end{aligned}$$

and applying the defining property of market time $d\tau_t = M_t dt$, together with the Dubins-Schwarz construction $dW_{\tau_t} = \sqrt{M_t} dB_t$, we see that

$$S_t = x_0 + \beta \int_0^t (1 - S_s) \, M_s \, \mathrm{d}s + \int_0^t \sqrt{S_s M_s} dB_s.$$

As a consequence, the bivariate stochastic process (S, M) has a t-time stochastic differential representation of the form

$$\begin{aligned} dS_t &= \beta(1 - S_t) M_t dt + \sqrt{S_t M_t} dB_t && (4.9) \\ dM_t &= a(M_t) dt + b(M_t) dB_t, && (4.10) \end{aligned}$$

where $S_0 = x_0$ and $M_0 \sim \mu$ with μ being the invariant distribution of (M_t) under $\mathbb{P}$.

Remark 4.2.4. A few comments regarding the construction of (X_t) seem to be in order. Firstly, subject to the Feller condition $\beta \geq 1/2$ in (2.1), the value of the asymptotic mean

$$\lim_{t\to\infty} \mathbb{E}\left(X_t \mid X_0 = x_0\right) = 1$$

has been chosen to match our particular choice of normalization; see Section 4.7.2 for an example of how we normalize using empirical observations of the S&P 500.

Furthermore, for a given parametric diffusion model $M_t = M_t(\gamma)$ where $\gamma \in \Gamma \subset \mathbb{R}^d$, the natural generalization

$$dX_t = \beta(1 - X_t)dt + \sigma\sqrt{X_t}dW_t, \tag{4.11}$$

imposes an over-parametrization on (S, M) in the sense that $\theta \mapsto \mathbb{P}_\theta$, where

$$\theta := (\beta, \sigma, \gamma) \in \Theta \subset \mathbb{R}_+^2 \times \mathbb{R}^d$$

is no longer an injection. To see this, we map $M_t \mapsto \tilde{M}_t = \sigma^2 M_t$, $\beta \mapsto \tilde{\beta} = \beta/\sigma^2$ and observe that

$$dS_t = \tilde{\beta}(1 - S_t)\tilde{M}_t dt + \sqrt{S_t\tilde{M}_t}dB_t.$$

Popular market activity models (M_t) that lead to this type of over-parametrization include the square-root process of Cox, Ingersoll, and Ross (1985), the GARCH(1,1) diffusion model in Nelson (1990) and the 3/2 diffusion in Ahn and Gao (1999). Hence, the bivariate representation (4.9)-(4.10) of (S, M) shows that our implicit choice of $\sigma = 1$ in (4.11) poses no restriction on the class of time-changed square-root diffusion processes considered in this paper.

4.3 Model implications

Before turning to statistical matters related to (S_t), we emphasize some of its key probabilistic properties in this section. Our main interests are model-implied volatility dynamics and leverage effects.

4.3.1 Volatility

To examine the volatility dynamics of (S_t), let $v_t = S_t M_t$ denote its spot volatility process derived in (4.9) and note that, by stochastic integration by parts,

$$v_t = S_0 M_0 + \int_0^t S_s dM_s + \int_0^t M_s dS_s + \langle S, M\rangle_t\,,$$

where $\langle\,\cdot\,,\,\cdot\,\rangle_t$ denotes the continuous quadratic covariation process and, in particular,

$$dv_t = \left(S_t b(M_t) + \sqrt{S_t M_t^3}\right) dB_t + dt\text{-term}. \tag{4.12}$$

Stochastic volatility models where the asset and the volatility are driven by the same, singular Brownian motion are known as *one-factor* models.

4.3.2 Continuous leverage effect

Empirical studies often indicate a strong negative correlation between asset returns and their volatility; see e.g. Christie (1982) or Bollerslev, Litvinova, and Tauchen (2006). This property is known in the financial literature as the *leverage effect* and is often modelled by assuming that the observable asset (S_t) and its underlying volatility process (v_t) are driven by correlated Brownian motions; see e.g. the seminal paper by Heston (1993).

To define a nonparametric measure of leverage effect that can be applied to most path-continuous continuous-time processes applied in finance, including the index models defined in Section 4.2, we work with a class of processes commonly used in high-frequency econometrics; see Aït-Sahalia and Jacod (2014).

Definition 4.3.1. A one-dimensional process (S_t) is said to be a *continuous Itô semimartingale* (w.r.t. $\mathscr{G}$) if (S_t) is $\mathscr{G}$-adapted and

$$S_t = S_0 + \int_0^t b_s \,\mathrm{d}s + \int_0^t \sqrt{v_s} dB_s,$$

where

- (b_t) is $\mathscr{G}$-progressive with $\int_0^t |b_s| \,\mathrm{d}s < \infty$,
- (v_t) is $\mathscr{G}$-progressive with $\int_0^t v_s \,\mathrm{d}s < \infty$,
- (B_t) is a Brownian motion w.r.t. $\mathscr{G}$.

The restriction to continuous Itô semimartingales enables us to characterize the leverage effect in terms of the instantaneous correlation

$$\mathrm{Cor}(X, Y)_t = \frac{d\langle X, Y\rangle_t / dt}{\sqrt{(d\langle X\rangle_t / dt)(d\langle Y\rangle_t / dt)}}, \tag{4.13}$$

well-defined for arbitrary continuous Itô semimartingales (X_t) and (Y_t). In particular, if we suppose that (S_t) is a stochastic volatility process

$$\begin{aligned} S_t &= S_0 + \int_0^t b_s \,\mathrm{d}s + \int_0^t \sqrt{v_s} dB_s \\ v_t &= v_0 + \int_0^t \tilde{b}_s \,\mathrm{d}s + \int_0^t \sqrt{\tilde{v}_s} d\tilde{B}_s, \end{aligned}$$

where (S_t) and the latent volatility process (v_t) are both continuous Itô semimartingales with $\mathrm{Cor}(B, \tilde{B})_t = \rho$ for some $\rho \in [-1, 1]$, then

$$\mathrm{Cor}(S, v)_t = \frac{\rho\sqrt{v_t}\sqrt{\tilde{v}_t}}{\sqrt{v_t \tilde{v}_t}} = \rho. \tag{4.14}$$

The instantaneous correlation (4.13) is defined in e.g. Mykland and Zhang (2012), with a recent extension to Itô semimartingales with jumps in terms of the continuous part of the integrated covariation $[X, Y]$ in Kalnina and Xiu (2017). The extension in Kalnina and Xiu (2017) provides a simple way to include both asset and volatility jumps into the dynamics of (S_t) without altering the defining property (4.14). We use this observation to construct a nonparametric estimator $\hat{\rho}_n$ of ρ in Section 4.5.

For the one-factor models (S_t) defined in Section 4.2,

$$\mathrm{Cor}(S,v)_t = \mathrm{sgn}\left(\sqrt{S_t M_t}\left[S_t b(M_t) + \sqrt{S_t M_t^3}\right]\right) \in \{-1, 1\} \tag{4.15}$$

corresponding to perfectly negative, respectively, positive correlation. Since the values of (S_t) and (M_t) are nonnegative by construction, it is clear from (4.15) that if $b(\,\cdot\,) > 0$, the model-implied leverage effect in the sense of (4.14) is $\rho = 1$. To identify conditions on the diffusion coefficient $b(\,\cdot\,)$ that ensure perfect negative correlation we note that

$$\mathrm{Cor}(S,v)_t = -1 \Longleftrightarrow S_t b(M_t) + \sqrt{S_t M_t^3} < 0,$$

which is equivalent to

$$b(M_t) < -\sqrt{\frac{M_t^3}{S_t}}. \tag{4.16}$$

Example 4.3.2. Consider the GARCH(1,1)-type diffusion model

$$dM_t = \kappa(\eta - M_t)dt - \xi M_t dB_t, \tag{4.17}$$

where $\kappa > 0$, $\eta > 0$ and, in particular, $\xi > 0$ to ensure that the diffusion coefficient $b(x) = -\xi \cdot x$ is strictly negative. After basic calculations, we see that the inequality (4.16) is satisfied if and only if

$$\xi > \sqrt{\frac{M_t}{S_t}}.$$

Example 4.3.3. If we instead consider a 3/2 diffusion model of the form

$$dM_t = \kappa M_t(\eta - M_t)dt - \xi M_t^{3/2} dB_t, \tag{4.18}$$

where $\kappa > 0$, $\eta > 0$ and $\xi > 0$, we see that the inequality (4.16) holds for

$$\xi > \frac{1}{\sqrt{S_t}}. \tag{4.19}$$

Since, in practice, $S_t \approx 1$ due to our preliminary normalization of the index, the parameter restriction (4.19) is satisfied whenever the market trajectory is sufficiently volatile. For the empirical data that we consider in Section 4.7, a value of $\xi > 2$ seems to be sufficient to ensure perfect negative correlation; see Fig. 4.16.

Remark 4.3.4. For classical parametric stochastic volatility models where the leverage effect is supposed constant, empirical studies of the S&P 500 indicate that $\rho \approx -0,75$; see e.g. Aït-Sahalia and Kimmel (2007). Roughly the same estimate was obtained by Kalnina and Xiu (2017) in a nonparametric framework with stochastic leverage in the sense of A. Veraart and L. Veraart (2012). A basic assumption in this paper is that diversified stock indices are driven by a single Brownian motion, corresponding to perfect *negative* leverage effect, i.e. $\rho = -1$. This is similar to the construction of the Heston-Nandi model discussed in e.g. Gatheral (2006).

4.4 Parameter estimation

In the following we propose a two-step method for estimating the mean-reversion parameter $\beta > 0$ in the dynamics of (S_t), as well as any parameter $\gamma \in \Gamma \subset \mathbb{R}^d$ that may appear in the latent market activity process $M_t = M_t(\gamma)$. We suppose we observe a discretization

$$S_{t_0}, \dots, S_{t_n}$$

where the observation times $\{t_i\}_{i=0}^n$ are deterministic and equidistant and we write $t_i = i\Delta$ for the appropriate $\Delta > 0$. In Section 4.4.1 we show how to filter a trajectory of (M_t) or τ using a Lamperti transformation. This enables us to estimate γ using standard methods for time-homogeneous diffusion processes. Section 4.4.2 deals with the estimation of β which, despite our ability to filter τ from the discretization of (S_t), is complicated due to dependence between the base process (X_t) and τ. Our proposed estimator exploits the common driving Brownian motion to construct an estimating equation which can be solved for β.

4.4.1 Market activity

As a tractable property of the index model, we can apply the explicitness of the Lamperti transform (see e.g. Aït-Sahalia (2002)) of the base process

$$dX_t = \beta\,(1 - X_t)\,dt + \sqrt{X_t}dW_t$$

to filter a trajectory of the market activity as follows: the transformed process $h(X_t)$, where

$$h(x) = \int_0^x y^{-1/2}\,\mathrm{d}y = 2\sqrt{x}, \tag{4.20}$$

is a unit diffusion with a representation

$$dh(X_t) = \frac{1}{\sqrt{X_t}}\left[\beta(1 - X_t) - \frac{1}{4}\right]dt + dW_t.$$

Therefore, the time-changed process

$$dh(S_t) = \frac{1}{\sqrt{S_t}}\left[\beta(1 - S_t) - \frac{1}{4}\right]M_t dt + dW_{\tau_t}$$

and, in particular,

$$\langle h(S)\rangle_t = \langle h(X)\rangle_{\tau_t} = \int_0^t M_s \, \mathrm{d}s = \tau_t.$$

This enables us to filter a trajectory of market time τ as

$$\widehat{\tau}_{t_i} = \sum_{j\leq i}\left(h(S_{t_j}) - h(S_{t_{j-1}})\right)^2 = 4\sum_{j\leq i}\left(\sqrt{S_{t_j}} - \sqrt{S_{t_{j-1}}}\right)^2 \tag{4.21}$$

for $\Delta > 0$ sufficiently small; see e.g. Theorem I.4.47, Jacod and Shiryaev (2003). Estimation for discretely observed integrated diffusion processes $\int_0^t M_s \, \mathrm{d}s$ where

$$dM_t = a(M_t;\gamma)dt + b(M_t;\gamma)dB_t$$

and $\gamma \in \Gamma \subset \mathbb{R}^d$ was considered in Chapter 3 as well as by Ditlevsen and M. Sørensen (2004) within the framework of prediction-based estimating functions. The general theory of prediction-based estimating functions was developed by M. Sørensen (2000, 2011) and, in particular, explicit prediction-based estimating functions are available for the class of Pearson diffusions defined in Forman and M. Sørensen (2008). Contrast estimators for drift and diffusion parameters of integrated diffusions were studied by Gloter (2006).

An alternative approach in the presence of high-frequency observations of (S_t) is to filter the instantaneous market activity $\{M_{t_i}\}_{i=0}^n$ by averaging over blocks containing $l_n \ll n$ neighbouring observations. The use of overlapping blocks has a smoothing effect on $\{\hat{M}_{t_i}\}_{i=0}^n$ and is frequently applied for filtering of spot volatility in high-frequency econometrics. The method only requires continuity of $t \mapsto M_t(\omega)$ through

$$M_{t_i}(\omega) = \lim_{h\to 0}\frac{1}{h}\int_{t_i}^{t_i+h} M_s(\omega)\, \mathrm{d}s = \lim_{h\to 0}\frac{1}{h}\left(\tau_{t_i+h}(\omega) - \tau_{t_i}(\omega)\right),$$

which yields the pointwise estimator

$$\hat{M}_{t_i} = \frac{1}{l_n\cdot\Delta}\left(\hat{\tau}_{t_i+l_n\cdot\Delta} - \hat{\tau}_{t_i}\right) \tag{4.22}$$

for $\hat{\tau}$ defined in (4.21). At this point, any of the numerous methods for estimation in stationary parametric diffusion models can be applied to estimate γ; see e.g. H. Sørensen (2004) for a survey of a selection of these methods.

4.4.2 Mean-reversion

Estimation of the mean-reversion parameter $\beta > 0$ poses a difficult problem. The basic idea in the following is to utilize the Dubins-Schwarz relation $dW_{\tau_t} = \sqrt{M_t}dB_t$ to derive an estimating equation for β based on Lamperti transformations. Specifically, we show that

$$d\begin{pmatrix} h_1(S_t) \\ h_2(M_t)\end{pmatrix} = \begin{pmatrix} +1 \\ -1\end{pmatrix} dW_{\tau_t} + dt\text{-term} \tag{4.23}$$

for suitable functions h_1 and h_2 which, in turn, enables us to estimate β by preliminary filtering and estimation of (M_t) as outlined in Section 4.4.1. We illustrate the estimation method for the 3/2 diffusion considered in Example 4.3.3 and discuss the modification to other diffusion models (M_t) at the end of the section.[2]

As shown in Section 4.4.1,

$$dh_1(S_t) = \frac{1}{\sqrt{S_t}}\left[\beta(1-S_t) - \frac{1}{4}\right] M_t dt + dW_{\tau_t} \tag{4.24}$$

for $h_1(x) = 2\sqrt{x}$, regardless of our choice of (M_t). Moreover, by Itô's formula applied to $h_2(x) = \xi^{-1}\log(x)$,

$$dh_2(M_t) = \left(\frac{\kappa\eta}{\xi} - \left(\frac{\kappa}{\xi} + \frac{\xi}{2}\right) M_t\right) dt - dW_{\tau_t}, \tag{4.25}$$

which establishes (4.23).

To construct an estimating equation for β, note that by (4.25) and the construction of h_2,

$$W_{\tau_{t_n}} = \frac{\kappa\eta}{\xi} t_n - \left(\frac{\kappa}{\xi} + \frac{\xi}{2}\right)\tau_{t_n} - \xi^{-1}\log\left(\frac{M_{t_n}}{M_0}\right),$$

and, similarly, it follows from (4.24) that

$$\begin{aligned}
W_{\tau_{t_n}} &= 2\left(\sqrt{S_{t_n}} - \sqrt{S_0}\right) - \int_0^{t_n} \frac{1}{\sqrt{S_s}}\left[\beta(1-S_s) - \frac{1}{4}\right] M_s \,\mathrm{d}s \\
&= 2\left(\sqrt{S_{t_n}} - \sqrt{S_0}\right) - \sum_{i=1}^{n}\int_{t_{i-1}}^{t_i} \frac{1}{\sqrt{S_s}}\left[\beta(1-S_s) - \frac{1}{4}\right] M_s \,\mathrm{d}s \\
&\approx 2\left(\sqrt{S_{t_n}} - \sqrt{S_0}\right) - \beta\Delta\sum_{i=1}^{n} \frac{\left(1 - S_{t_{i-1}}\right) M_{t_{i-1}}}{\sqrt{S_{t_{i-1}}}} + \frac{1}{4}\Delta\sum_{i=1}^{n}\frac{M_{t_{i-1}}}{\sqrt{S_{t_{i-1}}}}.
\end{aligned}$$

Equating the two yields the (approximate) estimating equation

$$\begin{gathered}
\frac{\kappa\eta}{\xi} t_n - \left(\frac{\kappa}{\xi} + \frac{\xi}{2}\right)\tau_{t_n} - \xi^{-1}\log\left(\frac{M_{t_n}}{M_0}\right) \approx \\
2\left(\sqrt{S_{t_n}} - \sqrt{S_0}\right) - \beta\Delta\sum_{i=1}^{n} \frac{\left(1 - S_{t_{i-1}}\right) M_{t_{i-1}}}{\sqrt{S_{t_{i-1}}}} + \frac{1}{4}\Delta\sum_{i=1}^{n}\frac{M_{t_{i-1}}}{\sqrt{S_{t_{i-1}}}},
\end{gathered} \tag{4.26}$$

which we solve for β using $\hat{\tau}_{t_n}$ given by (4.21), filtered values of $\{M_{t_i}\}_{i=0}^n$ obtained via (4.22) and an estimated value $\hat{\gamma}_n = (\hat{\kappa}_n, \hat{\eta}_n, \hat{\xi}_n)$ of the unknown parameter $\gamma \in \Gamma$ of the market activity process.

[2] A similar construction as (4.23) in terms of (B_t) is possible, however, the one-factor model (S, M) requires a different approach than the multivariate Lamperti approach in Aït-Sahalia (2008) for *reducible* diffusions.

Remark 4.4.1. The same estimation approach can be applied if (M_t) is either a square-root diffusion process

$$dM_t = \kappa\,(\eta - M_t)\,dt - \xi\sqrt{M_t}dB_t \tag{4.27}$$

or the GARCH(1,1) diffusion from Example 4.3.2. The only difference is that we need to replace $h_2(x) = \xi^{-1}\log(x)$ with $h_2(x) = \xi^{-1}x$ and $h_2(x) = 2\xi^{-1}\sqrt{x}$, respectively, to obtain a unit diffusion as in (4.25).

4.5 Testing the one-factor hypothesis

This section is devoted to the construction of a nonparametric test for determining whether or not a one-factor model is appropriate for modeling of diversified stock indices. We formulate the test for continuous stochastic volatility models of the general form

$$\begin{aligned} S_t &= S_0 + \int_0^t b_s\,\mathrm{d}s + \int_0^t \sqrt{v_s}dB_s \\ v_t &= v_0 + \int_0^t \tilde{b}_s\,\mathrm{d}s + \int_0^t \sqrt{\tilde{v}_s}d\tilde{B}_s, \end{aligned}$$

where both the discretely observed asset (S_t) and the volatility process (v_t) are assumed to be continuous Itô semimartingales with $\mathrm{Cor}(B,\tilde{B})_t = \rho$ for $\rho \in [-1,1]$. We are interested in the null hypothesis

$$\mathcal{H}_0 : \rho = -1, \tag{4.28}$$

corresponding to a one-factor model driven by (B_t), versus the general alternative $\mathcal{H}_1 : \rho \neq -1$. Since we are testing for a parameter value at the boundary, the derivation of an asymptotic error distribution for any particular estimator $\hat{\rho}_n$ poses a challenging problem and, as a simpler alternative, we simulate the empirical distribution of ρ in this paper.

For the construction of a nonparametric estimator $\hat{\rho}_n$ from discrete observations

$$S_{t_0}, \ldots, S_{t_n}$$

we follow Section 3 in Kalnina and Xiu (2017). However, whereas Kalnina and Xiu (2017) allow leverage to be stochastic, our (simpler) objective is to construct a global estimator $\hat{\rho}_n$ by averaging over all observations. We divide the construction into 4 steps, starting with filtering of the latent spot volatility using a local block average as in (4.22) and subsequently defining an estimator for the spot covariation $d\left\langle S,v\right\rangle_t/dt$, the spot volatility $d\left\langle S\right\rangle_t/dt$ and the spot volatility of volatility $d\left\langle v\right\rangle_t/dt$, respectively, to obtain an estimator for

$$\rho = \frac{d\left\langle S,v\right\rangle_t/dt}{\sqrt{(d\left\langle S\right\rangle_t/dt)(d\left\langle v\right\rangle_t/dt)}}$$

as discussed in Section 4.3.

STEP 1: To filter the local spot volatility, we fix $l_n \ll n$ and let

$$\hat{v}_{t_i} = \frac{1}{l_n \cdot \Delta}\left(\langle \hat{S}\rangle_{t_i + l_n \cdot \Delta} - \langle \hat{S}\rangle_{t_i}\right),$$

where

$$\langle \hat{S}\rangle_{t_j} = \sum_{k \leq j}(S_{t_k} - S_{t_{k-1}})^2,$$

which is often referred to as realized volatility in the econometrics literature; see e.g. Barndorff-Nielsen and Shephard (2002b).

STEP 2: To obtain an estimator for spot volatility at terminal time $t_n = n\Delta$, we let

$$v_n^* = \frac{1}{n \cdot \Delta}\langle \hat{S}\rangle_{t_n} = \frac{1}{n \cdot \Delta}\sum_{k \leq n}(S_{t_k} - S_{t_{k-1}})^2.$$

STEP 3: Our estimator for the spot volatility of the volatility process at terminal time is

$$\tilde{v}_n^* = \frac{3}{2 \cdot l_n \cdot n \cdot \Delta}\sum_{j=0}^{n-l_n}\left[\left(\hat{v}_{t_{j+l_n}} - \hat{v}_{t_j}\right)^2 - \frac{4}{l_n}\hat{v}_{t_j}^2\right]. \tag{4.29}$$

STEP 4: Finally, we construct an estimator for the spot covariation as

$$c_n^* = \frac{2}{l_n \cdot n \cdot \Delta}\sum_{j=0}^{n-l_n}\left[\left(S_{t_{j+l_n}} - S_{t_j}\right) \times \left(\hat{v}_{t_{j+l_n}} - \hat{v}_{t_j}\right)\right], \tag{4.30}$$

where the factor 2 has been added to accommodate the bias identified in Section 2.3 of Wang and Mykland (2014).[3]

Combining STEP 1-4 yields the nonparametric estimator

$$\hat{\rho}_n := \frac{c_n^*}{\sqrt{v_n^*}\sqrt{\tilde{v}_n^*}}. \tag{4.31}$$

The correction terms for $\tilde{v}_n^*$ in (4.29) and c_n^* in (4.30) are consistent with findings in Vetter (2015) and Wang and Mykland (2014), respectively, and are due to the preliminary estimation of volatility in STEP 1. In particular, the use of integrated covariation measures for discretely observed processes, see e.g. Barndorff-Nielsen and Shephard (2004a) or Zhang (2011), would lead to inconsistency of $\hat{\rho}_n$ as $\Delta \to 0$. The spot estimators defined in STEP 1-4 can be modified to encompass jumps in (S_t) using thresholding as in Mancini (2009) or Mancini, Mattiussi, and Renò (2015), or bipower variation; see e.g. Barndorff-Nielsen and Shephard (2004b, 2006).

[3] The factor 2 does not appear in the spot covariation estimator in Section 3 of Kalnina and Xiu (2017) who apply a subsequent bias-correction for their estimator of integrated stochastic leverage.

Remark 4.5.1. The estimator of ρ in (4.31) is rather crude. Compared to the construction in Kalnina and Xiu (2017), we avoid sub-dividing the dataset into non-overlapping blocks of length $l_n < L_n \ll n$ by considering the full set of observations (i.e. we choose $L_n = n$). A better option may be to proceed exactly as in Kalnina and Xiu (2017) and eventually define $\hat{\rho}_n$ as the average value obtained over all blocks of length L_n.

With the construction of $\hat{\rho}_n$, we obtain a simulation-based test for a one-factor hypothesis as follows:

ONE-FACTOR BOOTSTRAP TEST

1. *Evaluate* $\hat{\rho}_n$ *from the discretization* $S_{t_0}, \ldots, S_{t_n}$,
2. *Simulate* N *trajectories* $\left\{ S_{t_i}^{(j)} \right\}_{i=0}^{n}$ *under* $\mathcal{H}_0$ *and evaluate* $\hat{\rho}_n^{(j)}$,
3. *If* $\hat{\rho}_n$ *lies within the 95% quantile of the empirical distribution of* $\{\hat{\rho}_n^{(j)}\}_{j=1}^N$, *accept* $\mathcal{H}_0$.

Remark 4.5.2. Note that in applications with a parametric model $S_t = S_t(\theta)$ for $\theta \in \Theta \subset \mathbb{R}^d$, simulation under $\mathcal{H}_0$ requires preliminary estimation of θ.

4.6 Monte Carlo results

To study the behaviour of the proposed estimators and test statistic, we consider in this section the special case where (S_t) is given by the stochastic differential equation

$$\begin{aligned} dS_t &= \beta(1 - S_t)M_t dt + \sqrt{S_t M_t} dB_t \\ dM_t &= \kappa M_t(\eta - M_t)dt - \xi M_t^{3/2} dB_t. \end{aligned}$$

In Section 4.6.1 we outline the discretization scheme used for simulating the process, Section 4.6.2 is devoted to the estimation method and, finally, Section 4.6.3 examines the simulation-based one-factor test described in Section 4.5.

4.6.1 Path visualization

To simulate the time-changed process (S_t) we apply a truncated Milstein discretization of the one-factor representation of (S, M); see Chapter 10, Kloeden and Platen (1999). Letting $\Delta = t_i - t_{i-1}$ and $\Delta B_i = B_{t_i} - B_{t_{i-1}} \sim \mathcal{N}(0, \Delta)$, it follows that

$$\begin{aligned} \begin{pmatrix} S_{t_i} \\ M_{t_i} \end{pmatrix} &= \begin{pmatrix} S_{t_{i-1}} \\ M_{t_{i-1}} \end{pmatrix} + \begin{pmatrix} \beta\left(1 - S_{t_{i-1}}\right) M_{t_{i-1}} \\ \kappa M_{t_{i-1}}\left(\eta - M_{t_{i-1}}\right) \end{pmatrix} \Delta + \begin{pmatrix} \sqrt[*]{S_{t_{i-1}} M_{t_{i-1}}} \\ -\xi M_{t_{i-1}}^{3/2} \end{pmatrix} \Delta B_i \\ &+ \begin{pmatrix} \frac{1}{4} M_{t_{i-1}}\left(1 - \xi \sqrt[*]{S_{t_{i-1}}}\right) \\ \frac{3}{4}\xi^2 M_{t_{i-1}}^2 \end{pmatrix} \left(\Delta B_i^2 - \Delta\right), \end{aligned}$$

where $\sqrt[*]{\cdot}$ denotes a truncated version of the square-root function, i.e. $\sqrt[*]{x} = \sqrt{\max(0, x)}$.

In the following we fix $\beta = 1$, $\gamma = (\kappa, \eta, \xi) = (10, 7, 3)$ and consider the time interval $[0, 1]$ with a discretization step $\Delta \approx 9.51 \cdot 10^{-6}$, corresponding to 5-minute observations over a 1-year period. This yields a total of $105, 120$ observations with initial values $S_0 = 1$ and $M_0 \approx 4.83$, the latter being the stationary mean of (M_t). A sample trajectory of (S_t) is shown in Fig. 4.1.

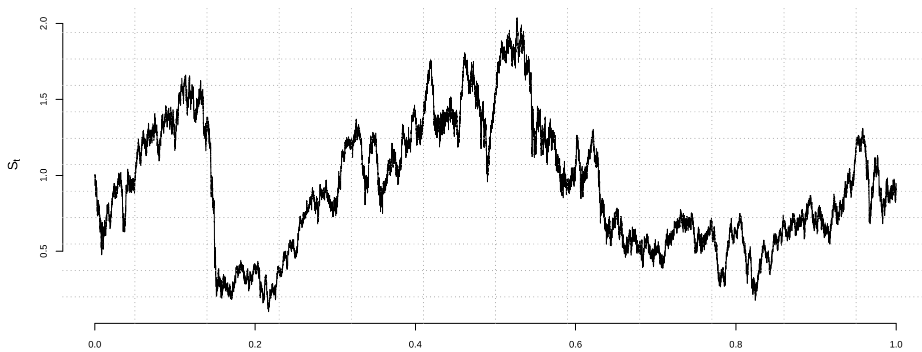

FIGURE 4.1: Monte Carlo discretization of (S_t) on $[0, 1]$.

From the discretization of (M_t), a simple approximation of instantaneous market time

$$\tau_{t_i} = \int_0^{t_i} M_s \, \mathrm{d}s = \sum_{j=1}^{i} \int_{t_{j-1}}^{t_j} M_s \, \mathrm{d}s \approx \Delta \sum_{j=1}^{i} M_{t_{j-1}} \tag{4.32}$$

and both processes are shown in Fig. 4.2 and Fig. 4.3, respectively. Of course, these discretizations are not available for estimation and testing purposes, however, as shown in Section 4.4.1, the explicitness of the Lamperti transform of $S_{\rho_t} = X_t$ enables us to filter out a trajectory of (M_t) from the discretization of (S_t).

Finally, we construct the latent spot volatility

$$v_{t_i} = S_{t_i} M_{t_i}$$

and plot the result in Fig. 4.4.

4.6.2 Estimation

In this section we study the behaviour of the two-step estimator described in Section 4.4. A preliminary undocumented simulation study of the 3/2 diffusion model reveals that long time series of 5-minute observations are necessary for reliable estimation of the drift parameters

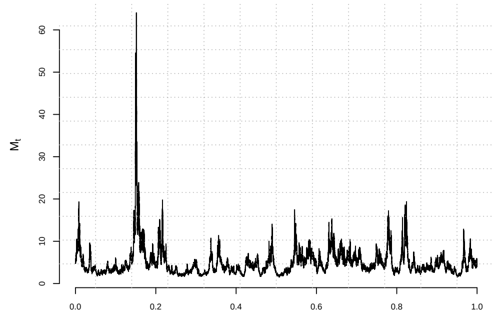

FIGURE 4.2: Monte Carlo discretization of 3/2 market activity process (M_t).

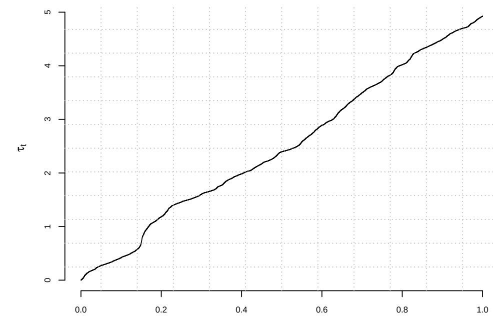

FIGURE 4.3: Approximate market time $\{\tau_{t_i}\}$ obtained from $\{M_{t_i}\}$ by (4.32).

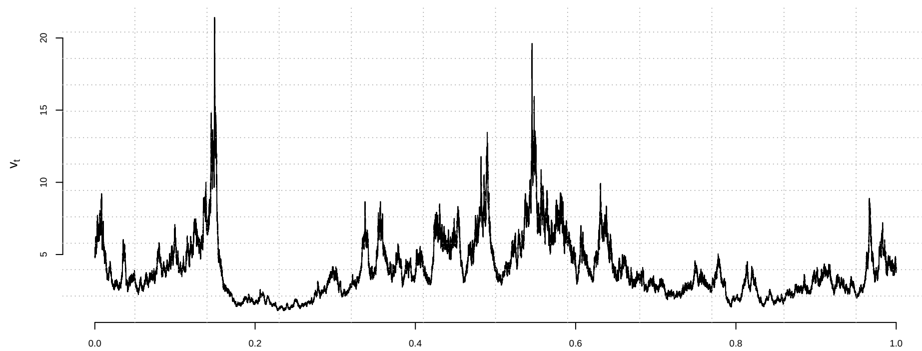

FIGURE 4.4: Discretization of the latent spot volatility $v_t = S_t M_t$.

κ and η. Consistent with our findings, we choose the fixed time horizon $[0, 5/2]$ in the following.

Given a discretization of (S_t), our first step is to filter out the market activity $\{\hat{M}_{t_i}\}_{i=0}^{n}$ using overlapping blocks of length $l_n \cdot \Delta$ of approximate market time; see (4.22). The outcome of a particular experiment is shown in Fig. 4.5.

For the parameter $\gamma = (\kappa, \eta, \xi)$ of the 3/2 diffusion (M_t), we apply the differential form

$$d \log(M_t) = \left(\kappa\eta - \left(\kappa + \frac{\xi^2}{2}\right) M_t\right) dt - \xi \sqrt{M_t} dB_t$$

to construct a natural estimator for ξ obtained by equating observed and theoretical quadratic variation as

$$\hat{\xi}_n^2 = \frac{\sum_{i=1}^{n} \left[\log\left(\hat{M}_{t_i}\right) - \log\left(\hat{M}_{t_{i-1}}\right)\right]^2}{\Delta \sum_{i=1}^{n} \hat{M}_{t_{i-1}}} \tag{4.33}$$

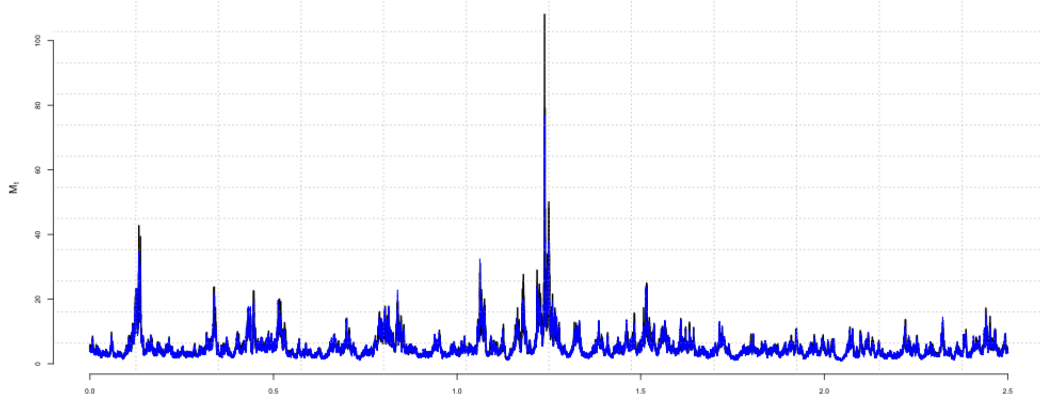

FIGURE 4.5: Time series of the true (simulated) market activity process $\{M_{t_i}\}$ (black) together with the filtered observations $\{\hat{M}_{t_i}\}$ (blue). As expected, the two trajectories follow each other very closely. The smoothing factor l_n is fixed at 100.

and, at this point, the explicit estimating functions defined in Section 4 of Kessler (2000) can be used to estimate κ and η; letting $g(x) = (x, \log(x))^T$ yields the two-dimensional estimating function

$$G_n(\gamma) = \sum_{i=1}^{n} \mathcal{L}_\gamma g\left(\hat{M}_{t_{i-1}}\right) = \sum_{i=1}^{n} \begin{pmatrix} \kappa \hat{M}_{t_{i-1}} \left(\eta - \hat{M}_{t_{i-1}}\right) \\ \kappa\left(\eta - \hat{M}_{t_{i-1}}\right) - \frac{1}{2}\hat{\xi}_n^2 \hat{M}_{t_{i-1}} \end{pmatrix},$$

where $\mathcal{L}_\gamma$ denotes the infinitesimal generator of (M_t). Hence, after solving the estimating equation $G_n(\gamma) = 0$, we obtain the explicit estimators

$$\begin{aligned} \hat{\eta}_n &= \frac{\frac{1}{n}\sum_{i=1}^{n} \hat{M}_{t_{i-1}}^2}{\frac{1}{n}\sum_{i=1}^{n} \hat{M}_{t_{i-1}}}, \\ \hat{\kappa}_n &= \frac{1}{2}\hat{\xi}_n^2 \left(\frac{\frac{1}{n}\sum_{i=1}^{n} \hat{M}_{t_{i-1}}^2}{\left(\frac{1}{n}\sum_{i=1}^{n} \hat{M}_{t_{i-1}}\right)^2} - 1 \right)^{-1}. \end{aligned}$$

Finally we solve the estimating equation (4.26) for β using estimated terminal market time $\hat{\tau}_T$ given by (4.21). Our findings are summarized in Table 4.6.[4]

[4]Note that estimation of the diffusion parameter ξ cannot be done using the estimating functions in Kessler (2000). As described in Section 5 of Hansen and Scheinkman (1995), simple moment equations can only be used to identify parameters in the invariant distribution of (M_t).

Table 4.6

Summary: two-step estimation of γ and β

β	$\hat{\tau}_T$	$\hat{\kappa}_n$	$\hat{\eta}_n$	$\hat{\xi}_n$	$\hat{\beta}_n$
1	12.25	9.59	7.18	2.99	0.90
1	11.87	11.21	6.72	3.05	1.36
1	11.81	9.90	6.96	3.06	1.42
2	11.98	11.33	6.75	3.05	2.21
2	11.94	11.08	6.76	3.04	2.57
2	12.36	10.73	7.00	2.99	1.54
3	12.33	11.71	6.82	2.99	2.35
3	12.00	8.92	7.29	3.04	3.75
3	12.11	11.67	6.74	3.02	3.13

Notes: (i) The true parameter $\gamma = (10, 7, 3)$. The values of the estimators $\hat{\gamma}_n$ and $\hat{\beta}_n$ are from particular realizations of (S_t) but suffice to illustrate convergence.

4.6.3 Testing the one-factor hypothesis

The index model (S_t) is a one-factor model by construction. In this section, we study the behaviour of the nonparametric test described in Section 4.5. To simulate the model under $\mathcal{H}_0$, we apply the estimated values $\hat{\gamma}_0 = (9.59, 7.18, 2.99)$ and $\hat{\beta}_0 = 0.90$ obtained in Table 4.6. Moreover, for the particular discretization that lead to $\hat{\gamma}_0$ and $\hat{\beta}_0$, the value of the test statistic $\hat{\rho}_n = -0.82$.

In the following, we simulate $N = 410$ trajectories $\{S_{t_i}^{(j)}\}_{i=0}^{n}$ on the time horizon $[0, 1]$ and in each case evaluate $\hat{\rho}_n^{(j)}$. The choice of time horizon $[0, 1]$ reduces computational time. As described in Step 1 of Section 4.5, the evaluation of $\hat{\rho}_n^{(j)}$ uses preliminary filtering of the spot volatility $\{v_{t_i}\}$ and a particular discretization $\{S_{t_i}^{(j)}\}$ is shown in Fig. 4.7 with a visual comparison of the latent volatility process and its filtered counterpart in Fig. 4.8. The outcome of the test is summarized in Fig. 4.9.

4.7 Empirical study: S&P 500

Whereas daily time series are available from numerous open source providers, access to research platforms is necessary for aggregation of high-frequency (intraday) financial data.[5]

[5]A standard open source provider is Yahoo! Finance (https://finance.yahoo.com).

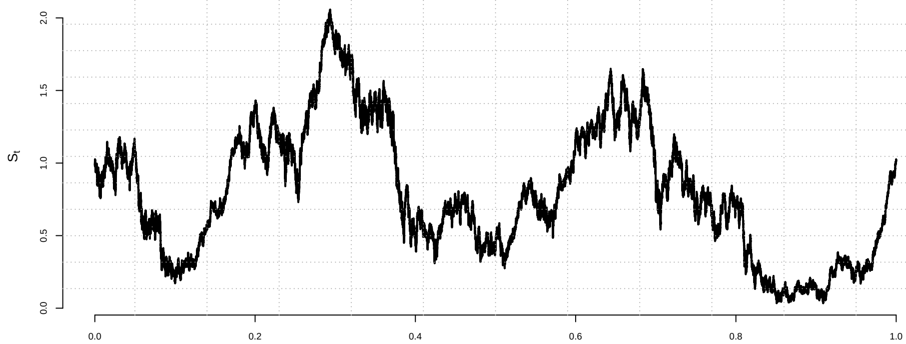

FIGURE 4.7: A Monte Carlo discretization of (S_t) on $[0, 1]$.

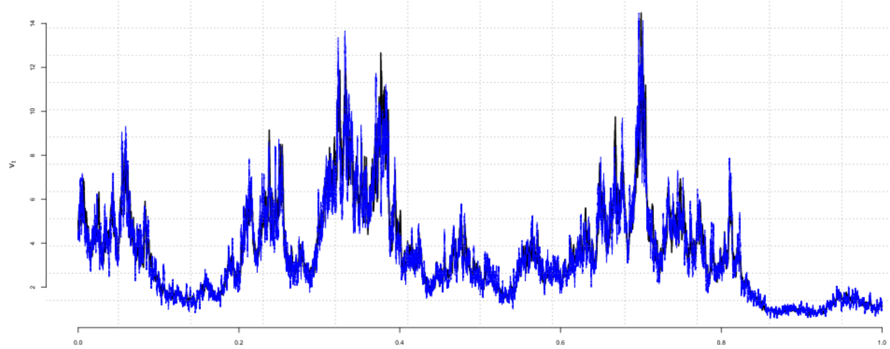

FIGURE 4.8: Time series of the true volatility process $\{v_{t_i}\}$ (black) together with the filtered observations $\{\hat{v}_{t_i}\}$ (blue). The smoothing factor l_n is fixed at 100.

The market data we apply in this section is pre-processed tick-by-tick transaction data of the SPDR S&P 500 ETF retrieved from Wharton Research Data Services. A detailed description of the pre-processing is given in Section 6.4 in Jönsson (2016) and consists of a two-step cleaning and subsequent downsampling to avoid irregular spacing. By downsampling to 5-minute observations, the effect of microstructure noise on the efficient value (price) of the index is negligible in practice; see e.g. Chapter 2, Aït-Sahalia and Jacod (2014). The down-sampled market data consists of $368,724$ price quotes and provides the basis for our empirical study.[6] We plot the data in Fig. 4.10.

[6]The SPDR S&P 500 ETF provides a liquid tradeable proxy of the S&P 500 Index and is fully suitable for our statistical purposes.

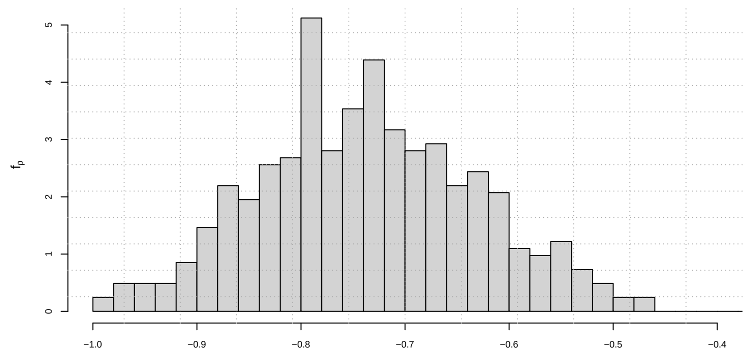

FIGURE 4.9: Empirical distribution of $\hat{\rho}_n$ under $\mathcal{H}_0 : \rho = -1$. For simulating the model, we apply the fixed set of null parameters $\hat{\gamma}_0 = (9.59, 7.18, 2.99)$ and $\hat{\beta}_0 = 0.90$. Of all the values of $\{\hat{\rho}_n^{(j)}\}_{j=1}^N$, two fell outside the sensible area $[-1, 1]$ and had to be rejected and resampled. The width of the empirical distribution can potentially be reduced by simulating longer trajectories which requires a substantial amount of computational time. As the value of the test statistic $\hat{\rho}_n = -0.82$ lies well within the 95% quantile of the empirical distribution, we cannot reject the one-factor hypothesis, as expected.

FIGURE 4.10: Log-transformed 5-minute observations of the SPDR S&P 500 ETF. The preprocessed time series consists of 368,724 price quotes spread uniformly across all trading days (9:30 - 16:00) from January 2, 1996, till December 31, 2013.

4.7.1 A closer look at the facts

As a preliminary step, we examine a few relevant qualitative properties of the SPDR S&P 500 ETF using nonparametric filtering methods. Specifically, we consider spot volatility, index jumps and the leverage effect.

Spot volatility To allow for jumps in the empirical index data, we use a local block average based on realized bipower variation to construct a daily measure of volatility, i.e. we let

$$\hat{v}_i = \frac{1}{h}\left(\mathrm{BV}(S)_{t_i+h} - \mathrm{BV}(S)_{t_i}\right), \quad i \geq 1$$

where t_i denotes the start of trading day $i-1$, h is the length of a trading day and

$$\mathrm{BV}(S)_{t_j} = \frac{\pi}{2}\sum_{k\leq j} |S_{t_k} - S_{t_{k-1}}| \cdot |S_{t_{k-1}} - S_{t_{k-2}}|.$$

The outcome is shown in Fig. 4.11.

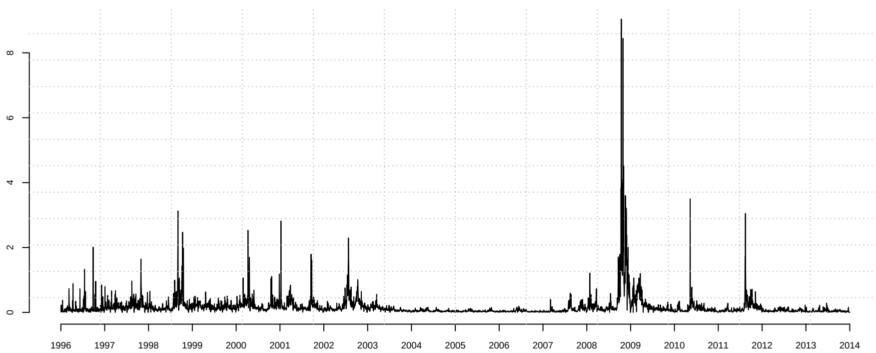

FIGURE 4.11: Filtered daily spot volatility $\{\hat{v}_i\}$ of the SPDR S&P 500 ETF.

Index jumps Our motivation for studying jump dynamics in the empirical data is two-fold. Firstly, the ability to remove index returns containing jumps shoud lead to better estimation of the time-changed index model (S_t) introduced in Section 4.2 and, secondly, Kalnina and Xiu (2017) argue that truncation of jumps stabilizes estimation of the continuous leverage effect. In this paper, we follow Lee and Mykland (2008) who propose a simple nonparametric test statistic for detecting infrequent jumps in high-frequency asset returns. The test is valid for all continuous Itô semimartingales with an independent jump component (J_t) and a summary of our findings for the SPDR S&P 500 ETF data is given in Fig. 4.12. Other nonparametric tests for the presence of asset jumps in high-frequency financial data include Barndorff-Nielsen and Shephard (2006) and Aït-Sahalia and Jacod (2009).

Continuous leverage effect To indicate the negative correlation between index returns and increments of the latent volatility process, we plot the normalized index (Section 4.7.2) together with the filtered daily spot volatility in Fig. 4.13.

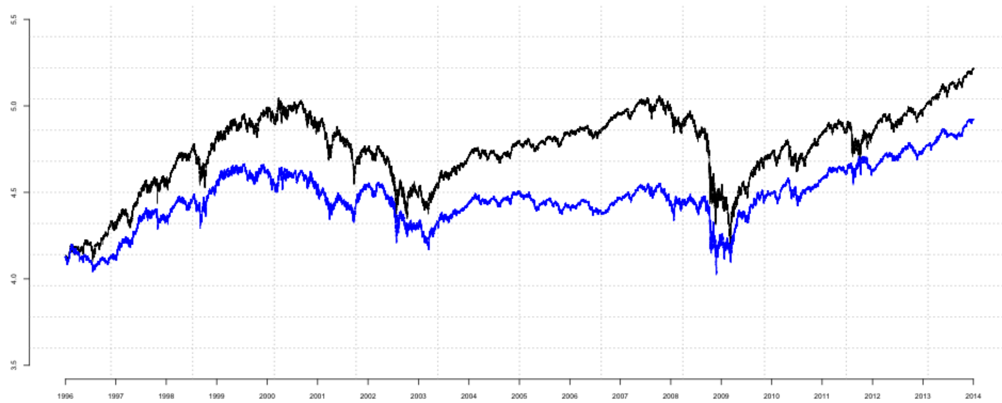

FIGURE 4.12: Visual comparison between the log-transformed SPDR S&P 500 ETF data (black) and a purely continuous component obtained by removing index returns containing jumps (blue). The jumps of the SPDR S&P 500 ETF have been removed using the test statistic proposed in Section 1 of Lee and Mykland (2008) with a significance level of 1%. The length of the rolling window used for preliminary estimation of spot volatility is $K = 270$.

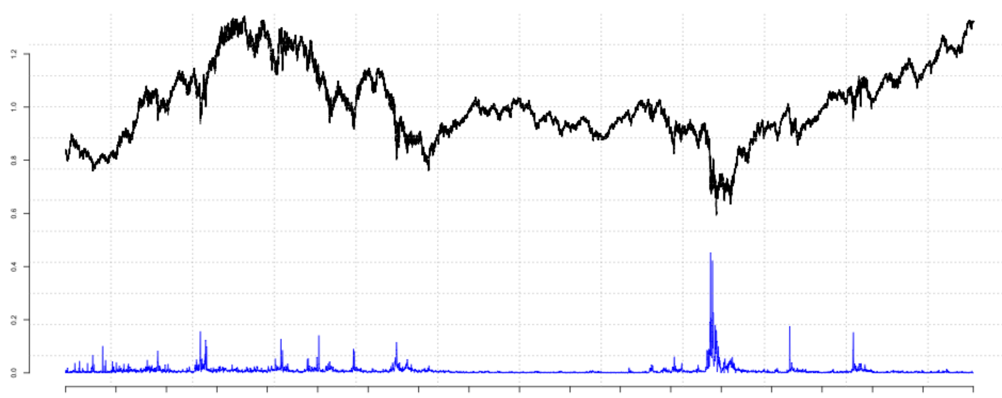

FIGURE 4.13: Normalized index data (black) together with the filtered trajectory of the spot volatility (blue). Visually, there seems to be an asymmetry between increments of the index and its volatility, corresponding to a negative leverage effect. The spot volatility has been reduced by a factor 20 for illustrative convenience.

4.7.2 Index normalization

As mentioned in Section 4.2, we normalize the index dynamics by subtracting a deterministic trend. The inclusion of a deterministic exponential trend is consistent with macroeconomic growth data displaying the evolution of US national GDP during the time period 1950-2017 (Fig. 4.14), as well as the log-transformed historical time series of daily closing values of the

S&P 500 for the same period (Fig. 4.15). The normalized time series is shown in Fig. 4.16.

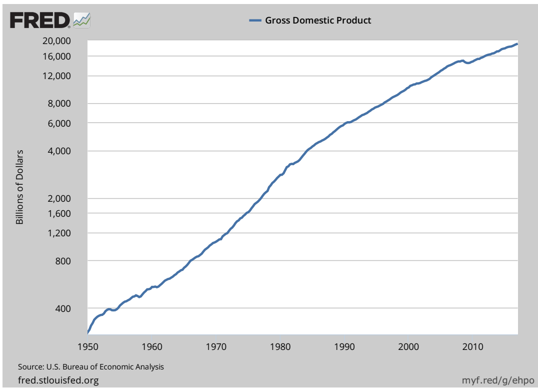

FIGURE 4.14: Log-scaled US national GDP, 1950-2017. Retrieved from FRED, Federal Reserve Bank of St. Louis; https://fred.stlouisfed.org/series/GDP, July 4, 2017.

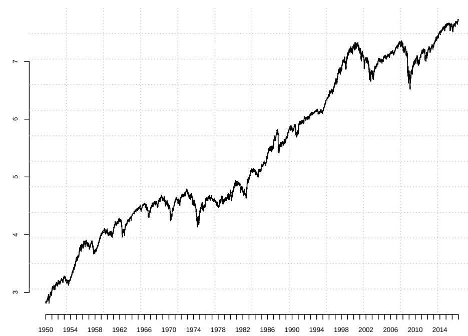

FIGURE 4.15: Log-transformed time series of daily closing values ($US) of the S&P 500 during the time period 03/01/1950 till 30/12/2016.

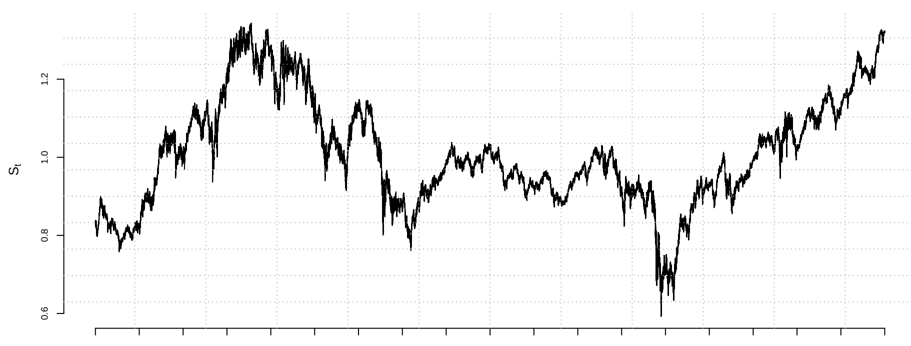

FIGURE 4.16: Time series of the normalized 5-minute data of the SPDR S&P 500 ETF. Before normalization, index returns containing jumps have been removed; see Section 4.7.1.

4.8 Extensions and concluding remarks

First and foremost, it remains to be seen how the estimators proposed in Section 4.4 and the one-factor test described in Section 4.5 behave on the empirical dataset. Another, more challenging, extension of apparent interest would be to derive the asymptotic distribution for the nonparametric leverage effect estimator $\hat{\rho}_n$ under $\mathcal{H}_0$ as $\Delta \to 0$. The simulation study in Section 4.6 seems to support our choice of bias-correction in Section 4.5, but a mathematical derivation would provide us with much needed clarification here.

Acknowledgments I am grateful to Eckhard Platen for hosting me at the UTS Business School, to Michael Sørensen for valuable discussions related to time-changed processes and proposing a simulation-based test for a one-factor hypothesis, to Per Mykland for stimulating conversations regarding estimation of leverage effect and, finally, Martin Jönsson for providing me with the pre-processed data of the SPDR S&P 500 ETF for use in the empirical study.

Bibliography

Ahn, D. and B. Gao (1999). “A Parametric Nonlinear Model of Term Structure Dynamics”. In: *Review of Financial Studies* 12.4, pp. 721–762.

Aït-Sahalia, Y. (1996a). “Nonparametric Pricing of Interest Rate Derivative Securities”. In: *Econometrica* 64.3, pp. 527–560.

— (1996b). “Testing Continuous-Time Models of the Spot Interest Rate”. In: *Review of Financial Studies* 9.2, pp. 385–426.

— (2002). “Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-form Approximation Approach”. In: *Econometrica* 70.1, pp. 223–262.

— (2008). “Closed-form likelihood expansions for multivariate diffusions”. In: *Annals of Statistics* 36.2, pp. 906–937.

Aït-Sahalia, Y. and J. Jacod (2009). “Testing for jumps in a discretely observed process”. In: *Annals of Statistics* 37.1, pp. 184–222.

— (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Aït-Sahalia, Y. and R. Kimmel (2007). “Maximum likelihood estimation of stochastic volatility models”. In: *Journal of Financial Economics* 83, pp. 413–452.

Aït-Sahalia, Y., P. Mykland, and L. Zhang (2005). “How Often to Sample a Continuous-Time Process in the Presence of Market Microstructure Noise”. In: *Review of Financial Studies* 18.2, pp. 351–416.

Ané, T. and H. Geman (2000). “Order Flow, Transaction Clock, and Normality of Asset Returns”. In: *Journal of Finance* 55.5, pp. 2259–2284.

Bainov, D. and P. Simeonov (1992). *Integral Inequalities and Applications*. Kluwer Academic Publishers.

Baltazar-Larios, F. and M. Sørensen (2010). “Maximum Likelihood Estimation for Integrated Diffusion Processes”. In: *Contemporary Quantitative Finance: Essays in Honour of Eckhard Platen*. Ed. by C. Chiarella and A. Novikov. Springer, pp. 407–423.

Bandi, F. and P. Phillips (2003). “Fully Nonparametric Estimation of Scalar Diffusion Models”. In: *Econometrica* 71.1, pp. 241–283.

Barndorff-Nielsen, O.E. (1997). “Normal Inverse Gaussian Distributions and Stochastic Volatility Modelling”. In: *Scandinavian Journal of Statistics* 24, pp. 1–13.

Barndorff-Nielsen, O.E. and N. Shephard (2002a). “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”. In: *Journal of the Royal Statistical Society* 64.2, pp. 253–280.

— (2002b). “Econometric analysis of realized volatility and its use in estimating stochastic volatility models”. In: *Journal of the Royal Statistical Society* 64.2, pp. 253–280.

Barndorff-Nielsen, O.E. and N. Shephard (2004a). "Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics". In: *Econometrica* 72.3, pp. 885–925.

— (2004b). "Power and Bipower Variation with Stochastic Volatility and Jumps". In: *Journal of Financial Econometrics* 2.1, pp. 1–37.

— (2006). "Econometrics of Testing for Jumps in Financial Economics Using Bipower Variation". In: *Journal of Financial Econometrics* 4.1, pp. 1–30.

Beskos, A., O. Papaspiliopoulos, and G. Roberts (2009). "Monte Carlo maximum likelihood estimation for discretely observed diffusion processes". In: *Annals of Statistics* 37.1, pp. 223–245.

Beskos, A., O. Papaspiliopoulos, G. Roberts, and P. Fearnhead (2006). "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion)". In: *Journal of the Royal Statistical Society* 68.3, pp. 333–382.

Bibby, B.M. and M. Sørensen (1995). "Martingale Estimation Functions for Discretely Observed Diffusion Processes". In: *Bernoulli* 1.1/2, pp. 17–39.

Bibinger, M., M. Jirak, and M. Vetter (2017). "Nonparametric change-point analysis of volatility". In: *Annals of Statistics* 45.4, pp. 1542–1578.

Bladt, M. and M. Sørensen (2014). "Simple simulation of diffusion bridges with application to likelihood inference for diffusions". In: *Bernoulli* 20.2, pp. 645–675.

Bollerslev, T., J. Litvinova, and G. Tauchen (2006). "Leverage and Volatility Feedback Effects in High-Frequency Data". In: *Journal of Financial Econometrics* 4.3, pp. 353–384.

Bollerslev, T. and H. Zhou (2002). "Estimating stochastic volatility diffusion using conditional moments of integrated volatility". In: *Journal of Econometrics* 109, pp. 33–65.

Broadie, M. and Ö. Kaya (2006). "Exact Simulation of Stochastic Volatility and Other Affine Jump Diffusion Processes". In: *Operations Research* 54.2, pp. 217–231.

Carr, P. et al. (2003). "Stochastic Volatility for Lévy Processes". In: *Mathematical Finance* 13.3, pp. 345–382.

Christie, A.A. (1982). "The Stochastic Behavior of Common Stock Variances: Value, Leverage and Interest Rate Effects". In: *Journal of Financial Economics* 10.4, pp. 407–432.

Clark, P.K. (1973). "A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices". In: *Econometrica* 41.1, pp. 135–155.

Comte, F., V. Genon-Catalot, and Y. Rozenholc (2007). "Penalized nonparametric mean square estimation of the coefficients of diffusion processes". In: *Bernoulli* 13.2, pp. 514–543.

— (2009). "Nonparametric adaptive estimation for integrated diffusions". In: *Stochastic Processes and their Applications* 119, pp. 811–834.

Cox, J., J. Ingersoll, and S. Ross (1985). "A Theory of the Term Structure of Interest Rates". In: *Econometrica* 53.2, pp. 385–407.

Dacunha-Castelle, D. and D. Florens-Zmirou (1986). "Estimation of the Coefficients of a Diffusion from Discrete Observations". In: *Stochastics* 19, pp. 263–284.

Ditlevsen, S. and M. Sørensen (2004). "Inference for Observations of Integrated Diffusion Processes". In: *Scandinavian Journal of Statistics* 31, pp. 417–429.

Doukhan, P. (1994). *Mixing: Properties and Examples.* Lecture Notes in Statistics 85. Springer-Verlag.

Drimus, G. (2012). "Options on realized variance by transform methods: a non-affine stochastic volatility model". In: *Quantitative Finance* 12.11, pp. 1679–1694.

Duffie, D. and K.J. Singleton (1993). "Simulated moments estimation of Markov models of asset prices". In: *Econometrica* 61.4, pp. 929–952.

Elerian, O., S. Chib, and N. Shephard (2001). "Likelihood Inference for Discretely Observed Nonlinear Diffusions". In: *Econometrica* 69.4, pp. 959–993.

Eraker, B. (2001). "MCMC Analysis of Diffusion Models With Application to Finance". In: *Journal of Business & Economic Statistics* 19.2, pp. 177–191.

Fan, J. (2005). "A Selective Overview of Nonparametric Methods in Financial Econometrics". In: *Statistical Science* 20.4, pp. 317–337.

Fan, J., Y. Li, and K. Yu (2012). "Vast Volatility Matrix Estimation Using High-Frequency Data for Portfolio Selection". In: *Journal of the American Statistical Association* 107.497, pp. 412–428.

Florens-Zmirou, D. (1989). "Approximate discrete-time schemes for statistics of diffusion processes". In: *Statistics* 20, pp. 547–557.

— (1993). "On estimating the diffusion coefficient from discrete observations". In: *Journal of Applied Probability* 30.4, pp. 790–804.

Forman, J.L. and M. Sørensen (2008). "The Pearson Diffusions: A Class of Statistically Tractable Diffusion Processes". In: *Scandinavian Journal of Statistics* 35, pp. 438–465.

Gatheral, J. (2006). *The Volatility Surface: A Practitioner's Guide*. Wiley.

Genon-Catalot, V. and J. Jacod (1993). "On the estimation of the diffusion coefficient for multi-dimensional diffusion processes". In: *Ann. Inst. Henri Poincaré* 29.1, pp. 119–151.

Genon-Catalot, V., T. Jeantheau, and C. Larédo (2000). "Stochastic volatility models as hidden Markov models and statistical applications". In: *Bernoulli* 6.6, pp. 1051–1079.

Genon-Catalot, V., C. Larédo, and D. Picard (1992). "Non-parametric Estimation of the Diffusion Coefficient by Wavelet Methods". In: *Scandinavian Journal of Statistics* 19.4, pp. 317–335.

Gloter, A. (2000). "Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient". In: *ESAIM: Probability and Statistics* 4, pp. 205–227.

— (2006). "Parameter Estimation for a Discretely Observed Integrated Diffusion Process". In: *Scandinavian Journal of Statistics* 33, pp. 83–104.

Gloter, A. and E. Gobet (2008). "LAMN property for hidden processes: The case of integrated diffusions". In: *Ann. Inst. Henri Poincaré* 44.1, pp. 104–128.

Gloter, A. and J. Jacod (2001a). "Diffusions with measurement errors. I. Local asymptotic normality". In: *ESAIM: Probability and Statistics* 5, pp. 225–242.

— (2001b). "Diffusions with measurement errors. II. Optimal estimators". In: *ESAIM: Probability and Statistics* 5, pp. 243–260.

Gobet, E., M. Hoffmann, and M. Reiß (2004). "Nonparametric estimation of scalar diffusions based on low frequency data". In: *Annals of Statistics* 32.5, pp. 2223–2253.

Godambe, V.P. and C.C. Heyde (1987). "Quasi-likelihood and Optimal Estimation". In: *International Statistical Review* 55.3, pp. 231–244.

Hansen, L.P. and J.A. Scheinkman (1995). "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes". In: *Econometrica* 63.4, pp. 767–804.

Hansen, L.P., J.A. Scheinkman, and N. Touzi (1998). "Spectral methods for identifying scalar diffusions". In: *Journal of Econometrics* 86.1, pp. 1–32.

Häusler, E. and H. Luschgy (2015). *Stable Convergence and Stable Limit Theorems*. Springer.

Heston, S.L. (1993). "A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options". In: *Review of Financial Studies* 6.2, pp. 327–343.

Heyde, C.C. (1997). *Quasi-Likelihood And Its Application: A General Approach to Optimal Parameter Estimation*. Springer.

Hoffmann, M. (1999a). "Adaptive estimation in diffusion processes". In: *Stochastic Processes and their Applications* 79.1, pp. 135–163.

— (1999b). "L_p estimation of the diffusion coefficient". In: *Bernoulli* 5.3, pp. 447–481.

Ignatieva, K. and E. Platen (2012). "Estimating the diffusion coefficient function for a diversified world stock index". In: *Computational Statistics & Data Analysis* 56.6, pp. 1333–1349.

Jacod, J. (2000). "Non-parametric Kernel Estimation of the Coefficient of a Diffusion". In: *Scandinavian Journal of Statistics* 27, pp. 83–96.

Jacod, J. and P. Protter (2012). *Discretization of Processes*. Springer-Verlag.

Jacod, J. and A.N. Shiryaev (2003). *Limit Theorems for Stochastic Processes*. Springer-Verlag.

Jakobsen, N.M. and M. Sørensen (2017). "Efficient estimation for diffusions sampled at high frequency over a fixed time interval". In: *Bernoulli* 23.3, pp. 1874–1910.

Jönsson, M. (2016). "Essays on Quantitative Finance". PhD thesis. University of Copenhagen.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer-Verlag.

Kalnina, I. and D. Xiu (2017). "Nonparametric Estimation of the Leverage Effect: A Trade-Off Between Robustness and Efficiency". In: *Journal of the American Statistical Association* 112.517, pp. 384–396.

Kessler, M. (1997). "Estimation of an Ergodic Diffusion from Discrete Observations". In: *Scandinavian Journal of Statistics* 24, pp. 211–229.

— (2000). "Simple and Explicit Estimating Functions for a Discretely Observed Diffusion Process". In: *Scandinavian Journal of Statistics* 27, pp. 65–82.

Kessler, M. and M. Sørensen (1999). "Estimating equations based on eigenfunctions for a discretely observed diffusion process". In: *Bernoulli* 5.2, pp. 299–314.

Kloeden, P. and E. Platen (1999). *Numerical Solution of Stochastic Differential Equations*. 3rd printing. Springer.

Le, T. and E. Platen (2006). "Approximating the growth optimal portfolio with a diversified world stock index". In: *Journal of Risk Finance* 7.5, pp. 559–574.

Lee, S. and P. Mykland (2008). "Jumps in Financial Markets: A New Nonparametric Test and Jump Dynamics". In: *Review of Financial Studies* 21.6, pp. 2535–2563.

Li, C. (2013). "Maximum-likelihood estimation for diffusion processes via closed-form density expansions". In: *Annals of Statistics* 41.3, pp. 1350–1380.

Li, J. and D. Xiu (2016). "Generalized Method of Integrated Moments for High-Frequency Data". In: *Econometrica* 84.4, pp. 1613–1633.

Madan, D.B. and E. Seneta (1990). "The Variance Gamma (V.G.) Model for Share Market Returns". In: *Journal of Business* 63.4, pp. 511–524.

Mancini, C. (2009). “Non-parametric Threshold Estimation for Models with Stochastic Diffusion Coefficient and Jumps”. In: *Scandinavian Journal of Statistics* 36, pp. 270–296.

Mancini, C., V. Mattiussi, and R. Renò (2015). “Spot volatility estimation using delta sequences”. In: *Finance and Stochastics* 19.2, pp. 261–293.

Mykland, P. and L. Zhang (2012). “The econometrics of high-frequency data”. In: *Statistical Methods for Stochastic Differential Equations*. Ed. by M. Kessler, A. Lindner, and M. Sørensen. CRC Press, pp. 109–190.

Nelson, D. (1990). “ARCH models as diffusion approximations”. In: *Journal of Econometrics* 45, pp. 7–38.

Pardoux, E. and A. Yu. Veretennikov (2001). “On the Poisson Equation and Diffusion Approximation. I”. In: *Annals of Probability* 29.3, pp. 1061–1085.

Platen, E. (2002). “Arbitrage in continuous complete markets”. In: *Advances in Applied Probability* 34.3, pp. 540–558.

— (2006). “A Benchmark Approach to Finance”. In: *Mathematical Finance* 16.1, pp. 131–151.

— (2011). “A Benchmark Approach to Investing and Pricing”. In: *The Kelly Capital Growth Investment Criterion*. Ed. by L.C. MacLean, E.O. Thorp, and W.T. Ziemba. World Scientific, pp. 409–426.

Platen, E. and D. Heath (2006). *A Benchmark Approach to Quantitative Finance*. Springer.

Platen, E. and R. Rendek (2012a). “Approximating the numéraire portfolio by naive diversification”. In: *Journal of Asset Management* 13.1, pp. 34–50.

— (2012b). *The Affine Nature of Aggregate Wealth Dynamics*. Research Paper 322. UTS Quantitative Finance Research Centre.

Renò, R. (2008). “Nonparametric Estimation of the Diffusion Coefficient of Stochastic Volatility Models”. In: *Econometric Theory* 24.5, pp. 1174–1206.

Revuz, D. and M. Yor (1999). *Continuous Martingales and Brownian Motion*. Springer-Verlag.

Roberts, G. and O. Stramer (2001). “On inference for partially observed nonlinear diffusion models using the Metropolis–Hastings algorithm”. In: *Biometrika* 88.3, pp. 603–621.

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill.

Sørensen, H. (2004). “Parametric Inference for Diffusion Processes Observed at Discrete Points in Time: a Survey”. In: *International Statistical Review* 72.3, pp. 337–354.

Sørensen, M. (2000). “Prediction-based estimating functions”. In: *Econometrics Journal* 3, pp. 123–147.

— (2011). “Prediction-based estimating functions: review and new developments”. In: *Brazilian Journal of Probability and Statistics* 25.3, pp. 362–391.

— (2012). “Estimating functions for diffusion-type processes”. In: *Statistical Methods for Stochastic Differential Equations*. Ed. by M. Kessler, A. Lindner, and M. Sørensen. CRC Press, pp. 1–107.

— (2017). “Efficient estimation for ergodic diffusions sampled at high frequency”. Working paper. URL: http://www.math.ku.dk/~michael/efficient.pdf.

Todorov, V. (2009). “Estimation of continuous-time stochastic volatility models with jumps using high-frequency data”. In: *Journal of Econometrics* 148.2, pp. 131–148.

Vasicek, O. (1977). “An equilibrium characterization of the term structure”. In: *Journal of Financial Economics* 5, pp. 177–188.

Veraart, A.E.D. and L.A.M. Veraart (2012). “Stochastic volatility and stochastic leverage”. In: *Annals of Finance* 8, pp. 205–233.

Veraart, A.E.D. and M. Winkel (2010). “Time Change”. In: *Encyclopedia of Quantitative Finance*. Ed. by R. Cont. Wiley, pp. 1–4.

Vetter, M. (2015). “Estimation of integrated volatility of volatility with applications to goodness-of-fit testing”. In: *Bernoulli* 21.4, pp. 2393–2418.

Wang, C. and P. Mykland (2014). “The Estimation of Leverage Effect With High-Frequency Data”. In: *Journal of the American Statistical Association* 109.505, pp. 197–215.

Yoshida, N. (1992). “Estimation for diffusion processes from discrete observation”. In: *Journal of Multivariate Analysis* 41.2, pp. 220–242.

Zhang, L. (2011). “Estimating covariation: Epps effect, microstructure noise”. In: *Journal of Econometrics* 160, pp. 33–47.