Prediction-based inference for integrated diffusions with high-frequency data

Emil S. Jørgensen and Michael Sørensen*
Dept. of Mathematical Sciences, University of Copenhagen
Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
E-mail: emil.joergensen@gmail.com and michael@math.ku.dk

November 25, 2025

Abstract

We consider parametric inference for an ergodic and stationary diffusion process, when the data are high-frequency observations of the integral of the diffusion process. Such data are obtained via certain measurement devices, or if positions are recorded and speed is modelled by a diffusion. In finance, realized volatility or variations thereof can be used to construct observations of the latent integrated volatility process. Specifically, we assume that the integrated process is observed at n equidistant, deterministic time points $i\Delta_n$ for some $\Delta_n > 0$ and consider the high-frequency/infinite horizon asymptotic scenario, where $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Subject to mild standard regularity conditions on (X_t), we prove the asymptotic existence and uniqueness of a consistent estimator for useful and tractable classes of prediction-based estimating functions. Asymptotic normality of the estimator is obtained under the additional rate assumption $n\Delta_n^2 \to 0$. The proofs are based on the useful Euler-Itô expansions of transformations of diffusions and integrated diffusions, which we study in some detail.

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

1 Introduction

Diffusion processes are used to model dynamical systems in many scientific areas, particularly in finance. While these processes are defined in terms of continuous-time dynamics, the available time series are observations of the system, or components of it, at discrete points in time. To bridge this gap between models and data, statistical methods for discretely observed continuous-time stochastic processes is a very active area of research, where the availability of high-frequency data has generated considerable interest in the construction and study of estimators and test statistics with nice asymptotic properties as the time between consecutive observations tends to zero.

*ORCID: 0000-0001-7233-5377

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

$$dI_t = X_t dt, \qquad I_0 = 0 \tag{1.1}$$

$$dX_t = a(X_t;\theta)dt + b(X_t;\theta)dB_t, \tag{1.2}$$

where the diffusion process (X_t) takes values in an open interval $(l, r) \subseteq \mathbb{R}$ and is ergodic with invariant distribution μ_θ. We assume that (X_t) is strictly stationary under the probability measure $\mathbb{P}_\theta$, i.e. that $X_0 \sim \mu_\theta$. The parameter θ takes values in $\Theta \subseteq \mathbb{R}^d$ for some $d \geq 1$.

Let the data be a single time series $\{I_{t_i^n}\}_{i=0}^n$ of observations of the integrated process at deterministic, equidistant points in time, i.e. $t_i^n = i\Delta_n$ for some $\Delta_n > 0$. The process (X_t) is latent. To enable consistent estimation of both drift and diffusion parameters, we consider the high-frequency/infinite horizon sampling scenario

$$n \to \infty, \quad \Delta_n \to 0, \quad n \cdot \Delta_n \to \infty, \tag{1.3}$$

where the time horizon tends to infinity with the number of observations. An equivalent observation scheme is given by the transformed variables

$$Y_i = \Delta_n^{-1}\left(I_{t_i^n} - I_{t_{i-1}^n}\right) = \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n} X_s \, \mathrm{d}s, \quad i = 1, \ldots, n. \tag{1.4}$$

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

We construct and study estimators using prediction-based estimating functions, which were proposed by Sørensen (2000, 2011) as a versatile framework for parametric inference in non-Markovian diffusion-type models. This approach was applied to integrated diffusions in Ditlevsen and Sørensen (2004). Their main contribution was to derive explicit Godambe-Heyde optimal prediction-based estimating functions for diffusions belonging to a tractable class of models that includes the Ornstein-Uhlenbeck process and the square-root (CIR) process and prove low-frequency asymptotic results. The main contribution of the present paper is to establish a high-frequency asymptotic theory for a class of prediction-based estimators, in particular, existence, uniqueness, consistency and asymptotic normality within the asymptotic scenario (1.3). Our proofs build on similar results for diffusion models in Jørgensen and Sørensen (2021).

Parametric estimation for discretely observed diffusion models (X_t) of the form (1.2) is the topic of numerous papers of which we can only list a few: Dacunha-Castelle and Florens-Zmirou (1986), Yoshida (1992), Hansen and Scheinkman (1995), Bibby and Sørensen (1995), Kessler (1997), Shoji and Ozaki (1998), Roberts and Stramer (2001), Aït-Sahalia (2002), Beskos *et al.* (2006), Bladt and Sørensen (2014), van der Meulen and Schauer (2017), Sørensen (2024), Pilipovic *et al.* (2024) and García-Portugués and Sørensen (2025), see also the review paper Sørensen (2012).

Although to a lesser extent, parametric inference for integrated diffusions has also been the topic of several papers in econometrics and statistics. In the econometric literature, the problem appears in the guise of continuous-time stochastic volatility models. To illustrate this, consider the simple stochastic volatility model for an asset price, $dS_t = \sqrt{v_t}dW_t$, where (W_t) denotes a standard Brownian motion. The availability of high-frequency observations of (S_t) enables us to filter out discrete time observations of the latent integrated volatility, $\int_0^t v_s \, \mathrm{d}s$, and view these as our data. Nonparametric filtering of integrated volatility from high-frequency time series is an emblematic problem in financial econometrics. An extensive list of references can be found in Aït-Sahalia and Jacod (2014). This procedure has lead to the construction of estimators for integrated processes in the case where the volatility dynamics are modeled by a time-homogeneous, stationary diffusion process similar to (1.2), e.g., 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 has been studied by e.g. Bollerslev and Zhou (2002), Barndorff-Nielsen and Shephard (2002) and Todorov (2009). Li and Xiu (2016) developed 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. Apart from the work by Ditlevsen and Sørensen (2004) that was summarized above, papers in the statistical literature include Baltazar-Larios and Sørensen (2010), who proposed a simulated EM-algorithm to obtain maximum likelihood estimators for integrated diffusions contaminated by noise, e.g. microstructure noise, and Gloter (2000, 2006), who proposed an approach that has significantly influenced the present paper. In this approach, which is based on expansion results for small values of Δ_n, the construction of contrast estimators utilizes that, as $\Delta_n \to 0$, $Y_i \approx X_{t_{i-1}^n}$, which allows high-frequency limit results for integrated diffusions to be established. Finally, nonparametric estimation of the drift and diffusion coefficient in the latent diffusion process from high-frequency observations of (I_t) was studied by Comte *et al.* (2009).

The paper is organized as follows. In Section 2, we present preliminaries: the notation and concepts used in the paper, our general assumptions on (X_t), and the prediction-based estimating functions considered in the paper. Section 3 contains an expansion of a transformation of the diffusion process of the form $f(X_{t_i^n}) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2} \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta)\varepsilon_{1,i} + \varepsilon_{2,i}$ and the similar result for $f(Y_i)$. The expansion for the integrated process, Y_i, was essentially pointed out by Gloter (2000). These expansions serve as essential building blocks for the asymptotic theory in our paper, and because they are related to the classic Euler approximation, we refer to them as *Euler-Itô expansions*. Section 4 is devoted to limit theorems for integrated diffusions, while the asymptotic results on existence, uniqueness, consistency and asymptotic normality of our estimators are developed in Section 5. Proofs and some auxiliary results are deferred to Section 6, and Section 7 concludes.

2 Preliminaries

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

2.1 Notation and concepts

Our general notation is as follows:

1. The true parameter value is denoted by θ_0.
2. We denote the state space of (X_t) by $(S, \mathscr{B}(S))$, where $S = (l, r)$ for $-\infty \leq l < r \leq \infty$ is an open interval equipped with its Borel σ-algebra $\mathscr{B}(S)$.
3. 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 functions f, for which $\mu_\theta(|f|^p) < \infty$. Moreover, $\mathscr{L}_0^p(\mu_\theta)$ denotes the subset of $\mathscr{L}^p(\mu_\theta)$ for which $\mu_\theta(f) = 0$.
4. By $\xrightarrow{\mathbb{P}_\theta}$ and $\xrightarrow{\mathscr{D}_\theta}$ we denote convergence in probability and in distribution under $\mathbb{P}_\theta$.

5. A function $f : S \times \Theta \to \mathbb{R}$ is said to be of *polynomial growth in* x if there exists a $C_\theta > 0$ such that $|f(x;\theta)| \leq C_\theta(1 + |x|^{C_\theta})$ for all $x \in S$.

6. In this paper, $R(\Delta, x; \theta)$ denotes a generic real function such that

$$|R(\Delta, x;\theta)| \leq F(x;\theta), \tag{2.1}$$

   where F is of polynomial growth in x.

7. For real functions f and g defined on a measure space $(A, \mathscr{A}, \nu)$, we write $f \leq_C g$ if there exists a constant $C > 0$ such that $f(a) \leq Cg(a)$, for ν-almost all $a \in A$. In particular, f and g can be random variables.

8. 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.

   We define $\mathcal{C}_p^j(S)$ analogously as a class of function $f : S \to \mathbb{R}$.

9. The *infinitesimal generator* of a diffusion process (X_t) is denoted by $\mathcal{A}_\theta$, and the corresponding domain by $\mathcal{D}_{\mathcal{A}_\theta}$. If (X_t) satisfies Condition 2.1 below, then $\mathcal{C}_p^2(S) \subseteq \mathcal{D}_{\mathcal{A}_\theta}$, and for all $f \in \mathcal{C}_p^2(S)$, $\mathcal{A}_\theta f = \mathcal{L}_\theta f$, where

$$\mathcal{L}_\theta f(x) = a(x;\theta)\partial_x f(x) + \frac{1}{2} b^2(x;\theta)\partial_x^2 f(x); \tag{2.2}$$

   see e.g. Kessler (2000).

10. For any diffusion process (X_t), the *potential* operator is given by

$$U_\theta(f)(x) = \int_0^\infty P_t^\theta f(x)\, \mathrm{d}t. \tag{2.3}$$

   It is defined for functions $f : S \to \mathbb{R}$ in the set $\mathcal{D}_{U_\theta} = \{f : \int_0^\infty |P_t^\theta f(x)|\, \mathrm{d}t < \infty\}$, where P_t^θ denotes the *transition operator* $P_t^\theta f(x) = \mathbb{E}_\theta\left(f(X_t) \mid X_0 = x\right)$.

11. We define

$$\mathscr{H}_\theta = \{f \in \mathcal{C}_p^4(S) \cap \mathcal{D}_{U_\theta} : \mu_\theta(f) = 0, U_\theta(f) \in \mathcal{C}_p^2(S)\}. \tag{2.4}$$

The potential operator plays an important role in our asymptotic theory. General results ensuring that $f \in \mathcal{D}_{U_\theta}$ and regularity of $U_\theta(f)$ can be found in Pardoux and Veretennikov (2001). For an ergodic diffusion with invariant measure μ_θ, $f \in \mathcal{D}_{U_\theta}$ must necessarily satisfy $\mu_\theta(f) = 0$. The reason why the potential operator is important in our theory is that under regularity conditions it satisfies the Poisson equation $\mathcal{L}_\theta(U_\theta(f)) = -f$. If (X_t) satisfies Condition 2.1 below, this is the case for $f \in \mathscr{H}_\theta$, see e.g. Proposition 3.3 in Jørgensen and Sørensen (2021).

2.2 Model assumption

To establish asymptotic results for integrated diffusions of the general form (1.1)-(1.2), we impose the following regularity conditions on (X_t).

Condition 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* $\mathcal{F}_t = \sigma\left(X_0, (B_s)_{s\leq t}\right)$, X_0 *is independent of* (B_t) *and*

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

Moreover, the 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* $k \geq 1$.

We define a discretized filtration by $\mathcal{F}_i^n := \mathcal{F}_{t_i^n}$.

Easily checked conditions for ρ-mixing of one-dimensional diffusion processes are given in Genon-Catalot *et al.* (2000). In particular, for an ergodic and time-reversible diffusion process, the ρ-mixing property is equivalent to the existence of a spectral gap. The latter means that the largest non-zero eigenvalue of the generator $\mathcal{A}_\theta$ of the diffusion process is strictly smaller that zero. From spectral theory it is known that all eigenvalues are non-positive. The size of the spectral gab, which we denote by λ_θ, equals minus the largest non-zero eigenvalue.

Under Condition 2.1, it is well-known that for $f \in \mathscr{L}_0^2(\mu_\theta)$ it holds that that $\|P_t^\theta f\|_2 \leq e^{-\lambda t}\|f\|_2$ for all $t \geq 0$, where $\|f\|_2 = \mu_\theta(f^2)^{\frac{1}{2}}$, see e.g. Lemma 3.2 in Jørgensen and Sørensen (2021). Using this we can define $\int_0^\infty P_t^\theta f(x)dt$ as the $\|\cdot\|_2$-limit of $\int_0^N P_t^\theta f(x)dt$ as $N \to \infty$. This limit exists and belongs to $\mathscr{L}_0^2(\mu_\theta)$ because $\int_0^N P_t^\theta f dt$ is a Cauchy sequence in $\mathscr{L}_0^2(\mu_\theta)$. Thus under Condition 2.1, U_θ is a well-defined mapping $\mathscr{L}_0^2(\mu_\theta) \mapsto \mathscr{L}_0^2(\mu_\theta)$, and since $\mathcal{C}_p^4(S) \subseteq \mathscr{L}^2(\mu_\theta)$ we have that $\mathscr{H}_\theta \subseteq \mathscr{L}_0^2(\mu_\theta) \subseteq D_{U_\theta}$. In particular, the space $\mathscr{H}_\theta$ can be written as

$$\mathscr{H}_\theta = \{f \in \mathcal{C}_p^4(S) : \mu_\theta(f) = 0, U_\theta(f) \in \mathcal{C}_p^2(S)\}. \tag{2.5}$$

The following condition on the true parameter value θ_0 is essential to the asymptotic theory for our estimators in Section 5. Here $\text{int}(\Theta)$ denotes the interior of Θ.

Condition 2.2. *The parameter space is* $\Theta \subseteq \mathbb{R}^d$ *and* $\theta_0 \in \text{int}(\Theta)$.

The notation $\mu_0 = \mu_{\theta_0}$, $\mathbb{P}_0 = \mathbb{P}_{\theta_0}$, etc., is applied throughout the paper.

2.3 Prediction-based estimating functions

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

$$G_n(\theta) = \sum_{i=q+1}^{n} \sum_{j=1}^{N} \pi_{i-1,j} \left[f_j(Y_i) - \breve{\pi}_{i-1,j}(\theta) \right] \tag{2.6}$$

where $\{f_j\}_{j=1}^N$ is a finite set of real-valued functions in $\mathscr{L}^2(\mu_\theta)$. 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

$$\mathcal{P}_{i-1,j} = \operatorname{span}\left\{1, f_j\left(Y_{i-1}\right), \ldots, f_j\left(Y_{i-q_j}\right)\right\} \subseteq \mathscr{L}^2(\mu_\theta), \tag{2.7}$$

where $q_j \geq 0$. The coefficients $\pi_{i-1,j}$ in (2.6) are d-dimensional column vectors with entries in $\mathcal{P}_{i-1,j}$, and $q := \max_{1 \leq j \leq N} q_j$.

The subspaces $\{\mathcal{P}_{i-1,j}\}_{ij}$ are called *predictor spaces*. What we predict are values of $f_j(Y_i)$ for $i \geq q+1$. Since every predictor space $\mathcal{P}_{i-1,j}$ is closed, the $\mathscr{L}^2(\mu_\theta)$-projection of $f_j(Y_i)$ onto $\mathcal{P}_{i-1,j}$, $\breve{\pi}_{i-1,j}(\theta)$, is well-defined and uniquely determined by the normal equations

$$\mathbb{E}_\theta \left(\pi \left[f_j(Y_i) - \breve{\pi}_{i-1,j}(\theta) \right] \right) = 0 \tag{2.8}$$

for all $\pi \in \mathcal{P}_{i-1,j}$. Moreover, by restricting our attention to a stationary process (X_t) and predictor spaces of the form (2.7), the solution to (2.8) is $\breve{\pi}_{i-1,j}(\theta) = \breve{a}_n(\theta)_j^T Z_{i-1,j}$, 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

$$\mathbb{E}_\theta \left[Z_{q_j,j} f_j(Y_{q_j+1}) \right] = \mathbb{E}_\theta \left[Z_{q_j,j} Z_{q_j,j}^T \right] \breve{a}_n(\theta)_j. \tag{2.9}$$

In the simplest case $q_j = 0$, $\mathcal{P}_{i-1,j} = \operatorname{span}\{1\}$ and, by (2.9), $\breve{\pi}_{i-1,j}(\theta) = \mathbb{E}_\theta f_j(Y_1)$.

We obtain an estimator $\hat{\theta}_n$ by solving the estimating equation $G_n(\theta) = 0$, and we call an estimator $\hat{\theta}_n$ a *G_n-estimator* if $\mathbb{P}_{\theta_0}(G_n(\hat{\theta}_n) = 0) \to 1$ as $n \to \infty$.

Most prediction-based estimating functions applied in practise are of the form considered here. In general, there is no explicit expression for the moments in (2.9). However, as noted by Ditlevsen and Sørensen (2004), polynomial functions $f_j(y) = y^{\beta_j}$, $\beta_j \in \mathbb{N}$, often enables calculation of the necessary moments by integrating over mixed moments of (X_t). This leads to explicit prediction-based estimating functions for the Pearson diffusions studied in Forman and Sørensen (2008).

3 Euler-Itô expansions

This section is devoted to expansions of transformations of diffusion processes and integrated diffusion processes observed over a small time interval of length Δ_n. We refer to these expansions as Euler-Itô expansions. Essentially, the following results provide a bridge between the asymptotic theory in Jørgensen and Sørensen (2021) and that of the present paper. The results are formulated with respect to an arbitrary probability measure $\mathbb{P}_\theta$.

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 3.1. *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*

$$f(X_{t_i^n}) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\partial_x f(X_{t_{i-1}^n})b(X_{t_{i-1}^n};\theta)\varepsilon_{1,i} + \varepsilon_{2,i}, \tag{3.1}$$

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

$$\begin{aligned} \mathbb{E}_\theta\left(\varepsilon_{2,i} \mid \mathcal{F}_{i-1}^n\right) &= \Delta_n \mathcal{L}_\theta f(X_{t_{i-1}^n}) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta), & (3.2)\\ \mathbb{E}_\theta\left(|\varepsilon_{2,i}|^k \mid \mathcal{F}_{i-1}^n\right) &= \Delta_n^k R(\Delta_n, X_{t_{i-1}^n};\theta), \quad k \geq 2. & (3.3) \end{aligned}$$

3.2 Integrated diffusions

To establish a similar result for functions of the integrated process, we rely on earlier work by Gloter (2000) as well as k'th order Taylor expansions of functions $f \in \mathcal{C}^k(S)$ of the form

$$f(Y_i) = \sum_{j=0}^{k-1} \frac{1}{j!}\partial_x^j f(X_{t_{i-1}^n})(Y_i - X_{t_{i-1}^n})^j + \frac{1}{k!}\partial_x^k f(Z_i^n)(Y_i - X_{t_{i-1}^n})^k, \tag{3.4}$$

where Z_i^n is 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 (3.4) for a given $k \geq 1$.

Lemma 3.2. *Let $h : S \to \mathbb{R}$ be of polynomial growth. Then, for any $k \geq 1$,*

$$\mathbb{E}_\theta\left(|h(Z_i^n)||(Y_i - X_{t_{i-1}^n})|^k \;\middle|\; \mathcal{F}_{i-1}^n\right) \leq_{C_k} \Delta_n^{k/2}(1 + |X_{t_{i-1}^n}|)^{C_k}. \tag{3.5}$$

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 3.2 implies that

$$\mathbb{E}_\theta\left(|f(Y_i) - f(X_{t_{i-1}^n})|^k \;\middle|\; \mathcal{F}_{i-1}^n\right) \leq_{C_k} \Delta_n^{k/2}(1 + |X_{t_{i-1}^n}|)^{C_k}. \tag{3.6}$$

Our main result in this section is of independent interest. It is a generalization of Proposition 2.2 in Gloter (2000). Note the resemblance with Proposition 3.1.

Proposition 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*

$$f(Y_i) = f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\partial_x f(X_{t_{i-1}^n})b(X_{t_{i-1}^n};\theta)\xi_{1,i} + \xi_{2,i}, \tag{3.7}$$

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

$$\begin{aligned} \mathbb{E}_\theta\left(\xi_{2,i} \mid \mathcal{F}_{i-1}^n\right) &= \Delta_n \mathcal{H}_\theta f(X_{t_{i-1}^n}) + \Delta_n^{3/2} R(\Delta_n, X_{t_{i-1}^n};\theta), & (3.8)\\ \mathbb{E}_\theta\left(\xi_{2,i}^2 \mid \mathcal{F}_{i-1}^n\right) &= \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta), & (3.9) \end{aligned}$$

with

$$\mathcal{H}_\theta f(x) = \frac{1}{2}\mathcal{L}_\theta f(x) - \frac{1}{12}b^2(x;\theta)\partial_x^2 f(x). \tag{3.10}$$

Moreover,

$$\mathbb{E}_\theta\left(\varepsilon_{1,i} \cdot \xi_{1,i}\right) = \frac{1}{2}, \tag{3.11}$$

where $\varepsilon_{1,i}$ is the random variable that appears in the Euler-Itô expansion (3.1).

4 Limit theory for integrated diffusions

As an application of the Euler-Itô expansion (3.7) and the corresponding bound (3.6), we derive in this section a law of large numbers and a central limit theorem for a class of functionals of integrated diffusions

$$\frac{1}{n}\sum_{i=1}^{n} f(Y_i), \tag{4.1}$$

where $f : S \to \mathbb{R}$ satisfies appropriate regularity conditions. For the remainder of the paper, all asymptotic results are obtained under the true probability measure $\mathbb{P}_0$ and under the asymptotic scenario (1.3).

Lemma 4.1. *Suppose that $f \in \mathcal{C}_p^1(S)$ and that (X_t) satisfies Condition 2.1. Then,*

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

The result of Lemma 4.1 appears in a slightly stronger version in Proposition 2 of Gloter (2006).

The result of the following lemma is that a central limit theorem for functionals (4.1) of integrated diffusions can be obtained under the same assumption on the rate of convergence of Δ_n and with the same Gaussian limit distribution as for similar functionals of discretely observed diffusion processes; see Proposition 3.4 in Jørgensen and Sørensen (2021).

Lemma 4.2. *Assume that $f \in \mathscr{H}_0$ and that (X_t) satisfies Condition 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

$$\mathcal{V}_0(f) = \mu_0\left(\left[\partial_x U_0(f) b(\,\cdot\,;\theta_0)\right]^2\right) = 2\mu_0\left(f U_0(f)\right). \tag{4.2}$$

The operator $U_0(f)$ appearing in the asymptotic variance (4.2) is the potential, which was defined and discussed in Subsections 2.1 and 2.2.

5 Asymptotic theory

This section contains our main asymptotic results on G_n-estimators obtained from prediction-based estimating functions of the type described in Subsection 2.3. The proofs are based on general asymptotic theory for estimating functions in Jacod and Sørensen (2018); see also Sørensen (2012). We confine the discussion to estimating functions of the form (2.6) where $N = 1$ and simplify the notation by writing

$$G_n(\theta) = \sum_{i=q+1}^{n} \pi_{i-1}\left[f(Y_i) - \breve{\pi}_{i-1}(\theta)\right], \tag{5.1}$$

$\mathcal{P}_{i-1}$ for the corresponding predictor spaces and so on for objects in Subsection 2.3 that depend on j. The extension to estimating functions with multiple predictor functions $\{f_j\}_{j=1}^{N}$ is discussed in Section 4.3 in Jørgensen and Sørensen (2021).

5.1 Simple predictor spaces

The simplest class of estimating functions of the form (5.1) occurs for $q = 0$. In this case, the orthogonal projection is $\breve{\pi}_{i-1}(\theta) = \mathbb{E}_\theta f(Y_1)$, and the one-dimensional predictor space $\mathcal{P}_{i-1}$ allows us to estimate one real parameter $\theta \in \Theta \subseteq \mathbb{R}$. Therefore, we consider the one-dimensional estimating function

$$G_n(\theta) = \sum_{i=1}^{n} \left[f(Y_i) - \mathbb{E}_\theta f(Y_1) \right]. \tag{5.2}$$

Similar estimating functions were studied for discretely observed diffusions by Kessler (2000).

Our study of the asymptotic properties of G_n-estimators is based on expansions of G_n in powers of Δ_n. In the simple case considered here, such an expansion follows easily from (3.7) in Proposition 3.3, which implies that for any $f \in \mathcal{C}_p^4(S)$

$$\mathbb{E}_\theta f(Y_1) = \mu_\theta(f) + \mathbb{E}_\theta(\xi_{2,1}) = \mu_\theta(f) + \Delta_n R(\Delta_n; \theta), \tag{5.3}$$

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

The following regularity conditions on G_n plus standard identifiability and rate conditions ensure existence, consistency and asymptotic normality of G_n-estimators.

Condition 5.1. *Suppose that*

- $f^*(x) := f(x) - \mu_0(f) \in \mathscr{H}_0$,
- $\theta \mapsto \mu_\theta(f) \in \mathcal{C}^1$,
- *For any compact subset $\mathcal{M} \subseteq \Theta$ and for Δ_n sufficiently small,*

$$\sup_{\theta \in \mathcal{M}} |\partial_\theta R(\Delta_n; \theta)| \leq C(\mathcal{M}). \tag{5.4}$$

Theorem 5.2. *Assume Conditions 2.1, 2.2 and 5.1 and the identifiability condition $\partial_\theta \mu_\theta(f) \neq 0$ for all $\theta \in \Theta$. Then the following assertions hold.*

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

$$\sqrt{n\Delta_n} \left(\hat{\theta}_n - \theta_0 \right) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, [\partial_\theta \mu_0(f)]^{-2} V_0(f)\right), \tag{5.5}$$

where $V_0(f) = 2\mu_0(f^* U_0(f^*))$.

Specifically, the statement about uniqueness means that for any G_n-estimator $\tilde{\theta}_n$ for which $\mathbb{P}_0(\tilde{\theta}_n \in \mathcal{K}) \to 1$, it holds that $\mathbb{P}_0(\hat{\theta}_n \neq \tilde{\theta}_n) \to 0$.

The identifiability condition and the assumption about the rate of convergence of Δ_n are exactly as in the similar result for prediction based estimating functions for discrete time observations of diffusion processes in Jørgensen and Sørensen (2021). Also the Gaussian the limit distribution is the same, which enables us to use the Monte Carlo method to calculate the asymptotic variance developed in Section 5.1 of Jørgensen and Sørensen (2021). Importantly, this method does not require an expression for the potential.

5.2 1-lag predictor spaces

The introduction of functions of past observations in the predictor space $\mathcal{P}_{i-1}$ increases the mathematical complexity considerably. Our main result establishes existence, uniqueness, consistency and asymptotically normality for prediction-based G_n-estimators with $q = 1$ under appropriate regularity conditions. In this case, the predictor space $\mathcal{P}_{i-1}$ is spanned by 1 and $f(Y_{i-1})$, and it follows from the normal equations (2.9) that the optimal predictor is

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

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

$$\breve{a}_n(\theta)_0 = \mathbb{E}_\theta f(Y_1)\left(1 - \breve{a}_n(\theta)_1\right), \tag{5.6}$$

$$\breve{a}_n(\theta)_1 = \frac{\mathbb{E}_\theta\left[f(Y_1)f(Y_2)\right] - \left[\mathbb{E}_\theta f(Y_1)\right]^2}{\mathbb{V}\mathrm{ar}_\theta f(Y_1)}. \tag{5.7}$$

Consistent with the two-dimensional predictor space, we consider $d = 2$ and investigate the estimating function

$$G_n(\theta) = \sum_{i=2}^{n} \begin{pmatrix} 1 \\ f(Y_{i-1}) \end{pmatrix} \left[f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1})\right] \tag{5.8}$$

for which the expansion in powers of Δ_n is more difficult than for (5.2).

Using on the Euler-Itô expansions in Section 3, we start by expanding the projection coefficients $\breve{a}_n(\theta)_0$ and $\breve{a}_n(\theta)_1$. As the proof is a bit long, we formulate the result in a separate lemma.

Lemma 5.3. *For $f \in \mathcal{C}_p^4(S)$, the projection coefficient vector $\breve{a}_n(\theta) = \left(\breve{a}_n(\theta)_0, \breve{a}_n(\theta)_1\right)^T$ has the expansion*

$$\breve{a}_n(\theta) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Delta_n \begin{pmatrix} -K_f(\theta)\mu_\theta(f) \\ K_f(\theta) \end{pmatrix} + \Delta_n^{3/2} R(\Delta_n;\theta) \tag{5.9}$$

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

$$K_f(\theta) = \mathbb{V}\mathrm{ar}_\theta f(X_0)^{-1}\left[\mu_\theta(f\mathcal{L}_\theta f) + \frac{1}{6}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right]. \tag{5.10}$$

The following regularity conditions on G_n are imposed in our asymptotic theory.

Condition 5.4. *Suppose that*

- $f_1^*(x) = K_f(\theta_0)\left[\mu_0(f) - f(x)\right] \in \mathscr{H}_0$,
- $f_2^*(x) = f(x)\mathcal{L}_0 f(x) + \frac{1}{6}[b(x;\theta_0)\partial_x f(x)]^2 - K_f(\theta_0)f(x)\left[f(x) - \mu_0(f)\right] \in \mathscr{H}_0$,
- $(\theta \mapsto \mu_\theta(f)) \in \mathcal{C}^1$, $(\theta \mapsto K_f(\theta)) \in \mathcal{C}^1$ *and the remainder term in (5.9) satisfies that*

$$\sup_{\theta\in\mathcal{M}} \|\partial_{\theta^T} R(\Delta_n;\theta)\| \leq C(\mathcal{M}), \tag{5.11}$$

for any compact subset $\mathcal{M} \subseteq \Theta$ and for Δ_n sufficiently small.

The matrix norm $\|\cdot\|$ in (5.11) and (5.13) can be chosen arbitrarily, and for convenience we suppose that $\|\cdot\|$ is submultiplicative. The following lemma establishes crucial technical steps in the proof of the main Theorem 5.6.

Lemma 5.5. *Assume that Conditions 2.1 and 5.4 holds. Then, for any $\theta \in \Theta$,*

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

where

$$\gamma(\theta_0;\theta) = \begin{pmatrix} K_f(\theta)(\mu_\theta - \mu_0)(f) \\ \mu_0(f\mathcal{L}_0 f) + \frac{1}{6}\mu_0\left([b(\,\cdot\,;\theta_0)\partial_x f]^2\right) - K_f(\theta)\left[\mu_0(f^2) - \mu_0(f)\mu_\theta(f)\right] \end{pmatrix}. \tag{5.12}$$

Moreover, for any compact subset $\mathcal{M} \subseteq \Theta$

$$\sup_{\theta\in\mathcal{M}} \left\| (n\Delta_n)^{-1} \partial_{\theta^T} G_n(\theta) - W(\theta) \right\| \xrightarrow{\mathbb{P}_0} 0, \tag{5.13}$$

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 5.6. *Assume Conditions 2.1, 2.2 and 5.4, that $W(\theta)$ is non-singular, and that the identifiability condition $\gamma(\theta_0;\theta) \neq 0$ for all $\theta \neq \theta_0$ is satisfied.*

Then the following assertions hold:

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

$$\sqrt{n\Delta_n}\left(\hat\theta_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2\left(0, \left[W(\theta_0)^{-1} V_0(f) (W(\theta_0)^{-1})^T\right]\right), \tag{5.14}$$

where

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

with

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

Compared to the results in Jørgensen and Sørensen (2021), the lower order $\Delta_n^{3/2}$ of the remainder term in the expansion (5.9) necessitates the rate assumption $n\Delta_n^2 \to 0$, which is stronger than what is needed for discretely observed diffusion processes. The same strong rate assumption appears in Gloter (2006) to ensure asymptotic normality for a class of minimum contrast estimators with observations of an integrated diffusion.

6 Proofs and auxiliary results

In this section we present the proofs of the results of the paper and some auxiliary results that are needed in the proofs.

6.1 Auxiliary results

We use several times that for a diffusion process X satisfying Condition 2.1 and a function $f \in \mathcal{C}_p^1(S)$, there exists, for every $k \geq 1$, a constant $C_{k,\theta} > 0$ such that

$$\mathbb{E}_\theta \left(\sup_{s\in[0,\Delta]} |f(X_{t+s}) - f(X_t)|^k \,\middle|\, \mathcal{F}_t \right) \leq C_{k,\theta} \Delta^{k/2} \left(1 + |X_t|\right)^{C_{k,\theta}}. \tag{6.1}$$

This classical result can be proved following the proofs of the similar results in Kessler (1997) and Gloter (2000).

We also use the well-known result that if $a(\cdot;\theta) \in \mathcal{C}_p^{2k,0}(S \times \Theta)$, $b(\cdot;\theta) \in \mathcal{C}_p^{2k,0}(S \times \Theta)$ and $f \in \mathcal{C}_p^{2(k+1)}(S)$ for a $k \geq 0$. Then

$$\mathbb{E}_\theta \left(f(X_{t+\Delta}) \mid \mathcal{F}_t \right) = \sum_{i=0}^{k} \frac{\Delta^i}{i!} \mathcal{L}_\theta^i f(X_t) + \Delta^{k+1} R(\Delta, X_t; \theta), \tag{6.2}$$

see e.g. Lemma 1.10 in Sørensen (2012).

Lemma 6.1. *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. Without loss of generality, we can assume that $t^* = 0$. Define $Z_t = \int_0^t H_s dX_s$. By stochastic integration-by-parts (the Itô- formula), $d(tZ_t) = tdZ_t + Z_t \,\mathrm{d}t$. Thus

$$\int_0^t t dZ_s = tZ_t = \int_0^t Z_s \,\mathrm{d}s + \int_0^t s dZ_s,$$

which verifies the result. □

6.2 Proofs

Since we study limits as $\Delta_n \to 0$, we can in all the proofs assume that Δ_n is bounded from above, e.g. $\Delta_n \leq 1$.

Proof of Proposition 3.1. By Itô's formula,

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

With the definitions

$$\begin{aligned}
\varepsilon_{1,i} &= \Delta_n^{-1/2} \int_{(i-1)\Delta_n}^{i\Delta_n} dB_s, & (6.3)\\
A_i &= \int_{(i-1)\Delta_n}^{i\Delta_n} \mathcal{L}_\theta f(X_s) \,\mathrm{d}s,\\
D_i &= \int_{(i-1)\Delta_n}^{i\Delta_n} \left[\partial_x f(X_s) b(X_s;\theta) - \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta) \right] dB_s,\\
\varepsilon_{2,i} &= A_i + D_i, & (6.4)
\end{aligned}$$

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

To prove the conditional moment expansions (3.2)-(3.3) apply Fubini's theorem followed by (6.2) to obtain

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

Furthermore, since $\mathbb{E}_\theta\left(\int_0^t [\partial_x f(X_s)b(X_s;\theta)]^2 \,\mathrm{d}s\right) = t\mu_\theta\left([f'b(\,\cdot\,;\theta)]^2\right) < \infty$, the stochastic integral $\int_0^t \partial_x f(X_s)b(X_s;\theta)dB_s$ is a $\mathbb{P}_\theta$-martingale, so $\mathbb{E}_\theta\left(D_i \mid \mathcal{F}_{i-1}^n\right) = 0$, which verifies (3.2).

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

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

and observe that, by Jensen's inequality,

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

Hence, by (6.1),

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

Similarly, with $h(x;\theta) = \partial_x f(x)b(x;\theta)$, the Burkholder-Davis-Gundy inequality (see e.g. Jacod and Protter (2012)), Jensen's inequality and (6.1) imply that for all $k \geq 2$,

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

and since $|\varepsilon_{2,i}|^k \leq_{C_k} |A_i|^k + |D_i|^k$, we conclude that $\mathbb{E}_\theta\left(|\varepsilon_{2,i}|^k \,\middle|\, \mathcal{F}^n_{i-1}\right) = \Delta_n^k R(\Delta_n, X_{t^n_{i-1}};\theta)$. □

Proof of Lemma 3.2. Since h is of polynomial growth,

$$|h(Z^n_i)| \leq_{C_k} 1 + |X_{t^n_{i-1}}|^{C_k} + |Y_i - X_{t^n_{i-1}}|^{C_k},$$

and by Jensen's inequality,

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

Hence the lemma follows because (6.1) implies that for any $k \geq 1$

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

□

Proof of Proposition 3.3. We start by proving the result for the identity mapping $f(x) = x$. In this case, $f' \equiv 1$ and $f'' \equiv 0$, so the Euler-Itô expansion (3.7) takes the form

$$Y_i = X_{t^n_{i-1}} + \Delta_n^{1/2} b(X_{t^n_{i-1}};\theta)\xi^*_{1,i} + \xi^*_{2,i}, \tag{6.5}$$

where asterisks $(*)$ are used to distinguish the remainder terms here from the general case. Here equation (3.8) has the form

$$\mathbb{E}_\theta\left(\xi^*_{2,i} \,\middle|\, \mathcal{F}^n_{i-1}\right) = \Delta_n \frac{1}{2} a(X_{t^n_{i-1}};\theta) + \Delta_n^{3/2} R(\Delta_n, X_{t^n_{i-1}};\theta). \tag{6.6}$$

By applying Lemma 6.1 to the stochastic integral, we find that

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

which, in turn, yields an expansion of the form (6.5) by defining

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

To verify the properties of $\xi^*_{1,i}$ and $\xi^*_{2,i}$, we observe that both are measurable w.r.t. $\mathcal{F}^n_i$, $\xi^*_{1,i}$ is Gaussian and independent of $\mathcal{F}^n_{i-1}$ and $\mathbb{E}_\theta(\xi^*_{1,i}) = 0$. Moreover, by Itô's isometry

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

To prove the conditional moment expansions of $\xi^*_{2,i}$, we first use the martingale property of $\int_0^t b(X_s;\theta)(i\Delta_n - s)dB_s$ to conclude that

$$\mathbb{E}_\theta\left(D_i \,\middle|\, \mathcal{F}^n_{i-1}\right) = \Delta_n^{-1}\cdot\mathbb{E}_\theta\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\left[b(X_s;\theta) - b(X_{t^n_{i-1}};\theta)\right](i\Delta_n - s)dB_s \,\middle|\, \mathcal{F}^n_{i-1}\right) = 0.$$

Therefore, $\mathbb{E}_\theta(\xi^*_{2,i}\,|\,\mathcal{F}^n_{i-1}) = \mathbb{E}_\theta(A_i\,|\,\mathcal{F}^n_{i-1})$. Application of Fubini's theorem and (6.2) shows that

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

and (6.6) follows because the last term equals $\Delta_n^2R(\Delta_n,x;\theta)$. In fact, we see that the slightly stronger result $\mathbb{E}_\theta(\xi^*_{2,i}\,|\,\mathcal{F}^n_{i-1}) = \Delta_n\frac{1}{2}a(X_{t^n_{i-1}};\theta) + \Delta_n^2R(\Delta_n,X_{t^n_{i-1}};\theta)$ holds for this particular choice of f.

To show that $\mathbb{E}_\theta((\xi^*_{2,i})^2\,|\,\mathcal{F}^n_{i-1}) = \Delta_n^2R(\Delta_n,X_{t^n_{i-1}};\theta)$, we use that by Jensen's inequality

$$|A_i|^2 \le \Delta_n^{-1}\int_{(i-1)\Delta_n}^{i\Delta_n}\left|\int_{(i-1)\Delta_n}^{s} a(X_u;\theta)\,\mathrm{d}u\right|^2\mathrm{d}s \le \sup_{s\in[0,\Delta_n]}\left|\int_{t^n_{i-1}}^{t^n_{i-1}+s} a(X_u;\theta)\,\mathrm{d}u\right|^2.$$

Moreover, for any $t\ge 0$ (again by Jensen's inequality),

$$\begin{aligned}
\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]}\left|\int_t^{t+s}a(X_u;\theta)\,\mathrm{d}u\right|^2\,\middle|\,\mathcal{F}_t\right) &\le \mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]} s\int_t^{t+s}|a(X_u;\theta)|^2\,\mathrm{d}u\,\middle|\,\mathcal{F}_t\right)\\
&= \Delta_n\mathbb{E}_\theta\left(\int_t^{t+\Delta_n}|a(X_u;\theta)|^2\,\mathrm{d}u\,\middle|\,\mathcal{F}_t\right).
\end{aligned}$$

Now by the linear growth of $a(\,\cdot\,;\theta)$ (Condition 2.1), $|a(X_u;\theta)|^2 \le_C 1 + |X_t|^2 + |X_u - X_t|^2$, so

$$\begin{aligned}
&\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]}\left|\int_t^{t+s}a(X_u;\theta)\,\mathrm{d}u\right|^2\,\middle|\,\mathcal{F}_t\right)\\
&\le_C \Delta_n^2(1+|X_t|^2) + \Delta_n\int_t^{t+\Delta_n}\mathbb{E}_\theta\left(|X_u - X_t|^2\,\middle|\,\mathcal{F}_t\right)\mathrm{d}u\\
&\le_C \Delta_n^2(1+|X_t|)^C + \Delta_n^2\cdot\mathbb{E}_\theta\left(\sup_{v\in[0,\Delta_n]}|X_{t+v} - X_t|^2\,\middle|\,\mathcal{F}_t\right) \le_C \Delta_n^2(1+|X_t|)^C,
\end{aligned}$$

where (6.1) implies the final inequality. In conclusion, $\mathbb{E}_\theta\left(|A_i|^2\,\middle|\,\mathcal{F}^n_{i-1}\right) \le_C \Delta_n^2(1+|X_{t^n_{i-1}}|)^C$. To obtain a similar bound for $|D_i|^2$, we apply the Burkholder-Davis-Gundy inequality, Jensen's inequality and (6.1) to obtain that

$$\begin{aligned}\mathbb{E}_\theta\left(|D_i|^2 \mid \mathcal{F}^n_{i-1}\right) &= \mathbb{E}_\theta\left(\Delta_n^{-2}\left|\int_{(i-1)\Delta_n}^{i\Delta_n}\left[b(X_s;\theta)-b(X_{t^n_{i-1}};\theta)\right](i\Delta_n-s)dB_s\right|^2 \,\middle|\, \mathcal{F}^n_{i-1}\right)\\ &\leq_C \mathbb{E}_\theta\left(\int_{(i-1)\Delta_n}^{i\Delta_n}\left[b(X_s;\theta)-b(X_{t^n_{i-1}};\theta)\right]^2 \mathrm{d}s \,\middle|\, \mathcal{F}^n_{i-1}\right)\\ &\leq \Delta_n\mathbb{E}_\theta\left(\sup_{s\in[0,\Delta_n]}\left|b(X_{t^n_{i-1}+s};\theta)-b(X_{t^n_{i-1}};\theta)\right|^2 \,\middle|\, \mathcal{F}^n_{i-1}\right)\\ &\leq_C \Delta_n^2(1+|X_{t^n_{i-1}}|)^C.\end{aligned}$$

and, as a consequence,

$$\mathbb{E}_\theta\left((\xi^*_{2,i})^2 \mid \mathcal{F}^n_{i-1}\right) = \Delta_n^2 R(\Delta_n, X_{t^n_{i-1}};\theta), \tag{6.7}$$

The extension to arbitrary $f \in \mathcal{C}^4_p(S)$ is based on Taylor expansions of the general form (3.4). First, a Taylor expansion combined with the Euler-Itô expansion (6.5), implies that

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

where

$$\xi_{1,i} = \xi^*_{1,i} = \Delta_n^{-3/2}\int_{(i-1)\Delta_n}^{i\Delta_n}(i\Delta_n - s)dB_s, \tag{6.8}$$

and $\xi_{2,i} = \sum_{k=1}^5 \xi^{(k)}_{2,i}$ with $\xi^{(1)}_{2,i} = \partial_x f(X_{t^n_{i-1}})\xi^*_{2,i}$, $\xi^{(2)}_{2,i} = \Delta_n\frac{1}{2}\partial_x^2 f(X_{t^n_{i-1}})b^2(X_{t^n_{i-1}};\theta)(\xi^*_{1,i})^2$, $\xi^{(3)}_{2,i} = \frac{1}{2}\partial_x^2 f(X_{t^n_{i-1}})(\xi^*_{2,i})^2$, $\xi^{(4)}_{2,i} = \Delta_n^{1/2}\partial_x^2 f(X_{t^n_{i-1}})b(X_{t^n_{i-1}};\theta)\xi^*_{1,i}\xi^*_{2,i}$ and $\xi^{(5)}_{2,i} = \frac{1}{6}\partial_x^3 f(Z^n_i)(Y_i - X_{t^n_{i-1}})^3$.

Each $\xi^{(k)}_{2,i}$, $k = 1, \ldots, 5$, is measurable w.r.t. $\mathcal{F}^n_i$ so it only remains to show that $\xi_{2,i}$ satisfies the moment expansions (3.8) and (3.9). By applying the previously derived conditional moment expansions $\mathbb{E}_\theta\left(\xi^*_{1,i} \mid \mathcal{F}^n_{i-1}\right) = 0$, $\mathbb{E}_\theta\left((\xi^*_{1,i})^2 \mid \mathcal{F}^n_{i-1}\right) = \frac{1}{3}$, (6.5) and (6.7) it follows immediately that

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

Furthermore, by Hölder's inequality,

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

implying $\mathbb{E}_\theta\left(\xi^{(4)}_{2,i} \,\middle|\, \mathcal{F}^n_{i-1}\right) = \Delta_n^{3/2}R(\Delta_n, X_{t^n_{i-1}};\theta)$, and finally, by Lemma 3.2, $\mathbb{E}_\theta\left(\xi^{(5)}_{2,i} \,\middle|\, \mathcal{F}^n_{i-1}\right) = \Delta_n^{3/2}R(\Delta_n, X_{t^n_{i-1}};\theta)$. Collecting our observations,

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

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 combine a lower order Taylor expansion with (6.5) to obtain

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

from which we get an alternative expression for the the remainder term $\xi_{2,i}$:

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

This expression 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, by applying (6.7) and Lemma 3.2, that $\mathbb{E}_\theta\left(\xi_{2,i}^2 \;\middle|\; \mathcal{F}_{i-1}^n\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta)$.

Finally, by the definitions (6.3) and (6.8) 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 4.1. The Lemma follows from Lemma 3.1 in Jørgensen and Sørensen (2021) if we show that

$$\frac{1}{n}\sum_{i=1}^n \left[f(Y_i) - f(X_{t_{i-1}^n})\right] = o_{\mathbb{P}_0}(1). \tag{6.9}$$

By applying the bound (3.6) for conditional expectations, we obtain

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

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

from which (6.9) follows by Lemma 9 in Genon-Catalot and Jacod (1993). □

Proof of Lemma 4.2. This result follows from Proposition 3.4 in Jørgensen and Sørensen (2021), if the following strengthening of (6.9) holds

$$\sqrt{n\Delta_n}\cdot\frac{1}{n}\sum_{i=1}^n \left[f(Y_i) - f(X_{t_{i-1}^n})\right] = o_{\mathbb{P}_0}(1), \tag{6.10}$$

To prove this, note that by Proposition 3.3,

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

where we use that $n\Delta_n^3 \to 0$. Moreover, the higher order bound (3.6) ensures that

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

and (6.10) follows from Lemma 9 in Genon-Catalot and Jacod (1993). □

Proof of Theorem 5.2. By applying the first order expansion (5.3) of $\mathbb{E}_\theta f(Y_1)$ together with Lemma 4.1, we see that

$$H_n(\theta) = \frac{1}{n}\sum_{i=1}^n [f(Y_i) - \mathbb{E}_\theta f(Y_1)] = \frac{1}{n}\sum_{i=1}^n [f(Y_i)-\mu_\theta(f)] + \Delta_n R(\Delta_n;\theta) \xrightarrow{\mathbb{P}_0} H(\theta),$$

where $H(\theta) = (\mu_0 - \mu_\theta)(f)$. Under Condition 5.1,

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

and for any compact subset $\mathcal{M}$ of Θ

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

Because $H(\theta_0) = 0$, we have now verified the conditions of Theorem 2.5 in Jacod and Sørensen (2018), from which the existence of a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ follows. That the estimator $\hat{\theta}_n$ is unique in any compact subset $\mathcal{K} \subseteq \Theta$ that contains θ_0 with $\mathbb{P}_0$-probability going to one as $n\to\infty$ follows from Theorem 2.7 in Jacod and Sørensen (2018), because the identifiability assumption implies that $H(\theta)\neq 0$ for $\theta \neq \theta_0$.

To establish asymptotic normality, note that (5.3), the additional rate assumption $n\Delta_n^3 \to 0$ and Lemma 4.2 ensure that

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

Now (5.5) follows by a standard Taylor expansion argument, see e.g. Theorem 2.11 in Jacod and Sørensen (2018). □

Proof of Lemma 5.3. We break the proof into three steps: **Step 1:** expand $[\mathbb{E}_\theta f(Y_1)]^2$, $\mathbb{E}_\theta f^2(Y_1)$ and $\mathbb{E}_\theta\left[f(Y_1)f(Y_2)\right]$ in powers of Δ_n, **Step 2:** eliminate $\mathcal{H}_\theta$ from the expansions, **Step 3:** calculate expansions of $\breve{a}_n(\theta)_0$ and $\breve{a}_n(\theta)_1$.

Step 1 Using Proposition 3.3, we find that

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

To expand $\mathbb{E}_\theta[f(Y_1)f(Y_2)]$ we note that by Proposition 3.3 this mixed moment equals

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

and then we expand the 9 terms of this expectation individually.

Term 1:

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

because by Proposition 3.1

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

Term 2:

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

because $\mathbb{E}_\theta\left(\xi_{1,2} \mid \mathcal{F}_{\Delta_n}\right) = 0$.

Term 3: By applying the moment expansions (3.8) and (6.2)

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

Term 4: By the Euler-Itô expansion (3.1),

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

The last equality holds because, by (3.11), $\mathbb{E}_\theta\left(\xi_{1,1}\varepsilon_{1,1} \mid \mathcal{F}_0\right) = 1/2$, and because by Hölder's inequality and (3.3), and since $\xi_{1,1} \sim \mathcal{N}(0, 1/3)$ and is independent of $\mathcal{F}_0^n$, we see that

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

Term 5: This term equals

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

since $\mathbb{E}_\theta\left(\xi_{1,2} \mid \mathcal{F}_{\Delta_n}\right) = 0$ by Proposition 3.3.

Term 6: This term equals

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

because by Hölder's inequality, $\left|\mathbb{E}_\theta\left(\xi_{1,1}\xi_{2,2} \mid \mathcal{F}_0\right)\right| \leq \mathbb{E}_\theta\left(\xi_{1,1}^2 \,\middle|\, \mathcal{F}_0\right)^{1/2} \mathbb{E}_\theta\left(\xi_{2,2}^2 \,\middle|\, \mathcal{F}_0\right)^{1/2}$, and by Proposition 3.3, $\mathbb{E}_\theta\left(\xi_{1,1}^2 \,\middle|\, \mathcal{F}_0\right) = 1/3$ and

$$\mathbb{E}_\theta\left(\xi_{2,2}^2 \,\middle|\, \mathcal{F}_0\right) = \mathbb{E}_\theta\left[\mathbb{E}_\theta\left(\xi_{2,2}^2 \,\middle|\, \mathcal{F}_{\Delta_n}\right) \,\middle|\, \mathcal{F}_0\right] = \mathbb{E}_\theta\left[\Delta_n^2 R(\Delta_n, X_{\Delta_n};\theta) \,\middle|\, \mathcal{F}_0\right] = \Delta_n^2 R(\Delta_n, X_0;\theta).$$

Term 7: By the Euler-Itô expansion (3.1),

$$\begin{aligned}&\mathbb{E}_\theta\left[f(X_{\Delta_n})\xi_{2,1}\right]\\ &= \mathbb{E}_\theta\left[f(X_0)\mathbb{E}_\theta\left(\xi_{2,1} \mid \mathcal{F}_0\right)\right] + \Delta_n^{1/2}\mathbb{E}_\theta\left[\partial_x f(X_0) b(X_0;\theta)\mathbb{E}_\theta\left(\varepsilon_{1,1}\xi_{2,1} \mid \mathcal{F}_0\right)\right] + \mathbb{E}_\theta(\varepsilon_{2,1}\xi_{2,1})\\ &= \Delta_n \mu_\theta(f\mathcal{H}_\theta f) + \Delta_n^{3/2} R(\Delta_n;\theta)\end{aligned}$$

where the last equality holds because, by Proposition 3.3, $\mathbb{E}_\theta\left(\xi_{2,1} \mid \mathcal{F}_0\right) = \Delta_n \mathcal{H}_\theta f(X_0) + \Delta_n^{3/2} R(\Delta_n, X_0;\theta)$, and by Hölder's inequality and Propositions 3.1 and 3.3, $\mathbb{E}_\theta\left(\varepsilon_{1,1}\xi_{2,1} \mid \mathcal{F}_0\right) = \Delta_n R(\Delta_n, X_0;\theta)$ and $\mathbb{E}_\theta(\varepsilon_{2,1}\xi_{2,1}) = \Delta_n^2 R(\Delta_n;\theta)$.

Term 8: By Proposition 3.3, this term equals

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

Term 9: By combining Hölder's inequality and (3.9), we obtain

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

Finally, we add the expansions of the nine terms and conclude that

$$\begin{aligned}&\mathbb{E}_\theta f(Y_1)f(Y_2) = \\ &\quad \mu_\theta(f^2) + \Delta_n\left(\mu_\theta(f\mathcal{L}_\theta f) + 2\mu_\theta(f\mathcal{H}_\theta f) + \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right)\right) + \Delta_n^{3/2}R(\Delta_n;\theta).\end{aligned}$$

Step 2 To eliminate $\mathcal{H}_\theta$ from the expansions of $[\mathbb{E}_\theta f(Y_1)]^2$, $\mathbb{E}_\theta f^2(Y_1)$ and $\mathbb{E}_\theta\left[f(Y_1)f(Y_2)\right]$, we rewite $\mu_\theta(\mathcal{H}_\theta f)$, $\mu_\theta(f\mathcal{H}_\theta f)$ and $\mu_\theta(\mathcal{H}_\theta f^2)$ using the definition of $\mathcal{H}_\theta$, (3.10), and that $\mu_\theta(\mathcal{L}_\theta f) = 0$ for all $f \in \mathcal{D}_{\mathcal{A}_\theta}$; see e.g. Hansen and Scheinkman (1995). It follows immediately that

$$\begin{aligned}\mu_\theta(\mathcal{H}_\theta f) &= -\frac{1}{12}\mu_\theta\left(b^2(\,\cdot\,;\theta)\partial_x^2 f\right) \\ \mu_\theta(f\mathcal{H}_\theta f) &= \frac{1}{2}\mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{12}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right).\end{aligned}$$

Moreover, since $\partial_x f^2 = 2f\partial_x f$ and $\partial_x^2 f^2 = 2\left[f\partial_x^2 f + (\partial_x f)^2\right]$,

$$\begin{aligned}\mathcal{H}_\theta f^2(x) &= \frac{1}{2}a(x;\theta)\partial_x f^2(x) + \frac{1}{6}b^2(x;\theta)\partial_x^2 f^2(x) \\ &= f(x)a(x;\theta)\partial_x f(x) + \frac{1}{3}f(x)b^2(x;\theta)\partial_x^2 f(x) + \frac{1}{3}[b(x;\theta)\partial_x f(x)]^2 \\ &= f(x)\mathcal{L}_\theta f(x) - \frac{1}{6}f(x)b^2(x;\theta)\partial_x^2 f(x) + \frac{1}{3}[b(x;\theta)\partial_x f(x)]^2,\end{aligned}$$

which shows that

$$\mu_\theta(\mathcal{H}_\theta f^2) = \mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{3}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right).$$

Thus

$$\begin{aligned}[\mathbb{E}_\theta f(Y_1)]^2 &= \mu_\theta(f)^2 + \Delta_n M_0(\theta) + \Delta_n^{3/2}R(\Delta_n;\theta) &\quad (6.11)\\ \mathbb{E}_\theta f^2(Y_1) &= \mu_\theta(f^2) + \Delta_n M_1(\theta) + \Delta_n^{3/2}R(\Delta_n;\theta) &\quad (6.12)\\ \mathbb{E}_\theta f(Y_1)f(Y_2) &= \mu_\theta(f^2) + \Delta_n M_2(\theta) + \Delta_n^{3/2}R(\Delta_n;\theta), &\quad (6.13)\end{aligned}$$

where

$$\begin{aligned}M_0(\theta) &= -\frac{1}{6}\mu_\theta(f)\mu_\theta\left(b^2(\,\cdot\,;\theta)\partial_x^2 f\right) \\ M_1(\theta) &= \mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{3}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right) \\ M_2(\theta) &= 2\mu_\theta(f\mathcal{L}_\theta f) - \frac{1}{6}\mu_\theta\left(f b^2(\,\cdot\,;\theta)\partial_x^2 f\right) + \frac{1}{2}\mu_\theta\left([b(\,\cdot\,;\theta)\partial_x f]^2\right).\end{aligned}$$

Step 3 From the moment expansions (6.11)-(6.13), it follows that

$$\begin{aligned}\breve{a}_n(\theta)_1 &= \frac{\mathbb{E}_\theta f(Y_1)f(Y_2) - [\mathbb{E}_\theta f(Y_1)]^2}{\mathbb{V}\mathrm{ar}_\theta f(Y_1)} \\ &= \frac{1 + \Delta_n \mathbb{V}\mathrm{ar}_\theta f(X_0)^{-1}(M_2(\theta) - M_0(\theta)) + \Delta_n^{3/2}R(\Delta_n;\theta)}{1 + \Delta_n \mathbb{V}\mathrm{ar}_\theta f(X_0)^{-1}(M_1(\theta) - M_0(\theta)) + \Delta_n^{3/2}R(\Delta_n;\theta)},\end{aligned}$$

and since $1/(1+x) = 1 - x + O(x^2)$, we obtain the expansion

$$
\begin{aligned}
\breve{a}_n(\theta)_1 &= 1 + \Delta_n \mathbb{V}\mathrm{ar}_\theta f(X_0)^{-1}\left[M_2(\theta) - M_1(\theta)\right] + \Delta_n^{3/2} R(\Delta_n;\theta) \\
&= 1 + \Delta_n K_f(\theta) + \Delta_n^{3/2} R(\Delta_n;\theta), \qquad (6.14)
\end{aligned}
$$

where $K_f(\theta)$ is given by (5.10).

Finally, since by (3.7) $\mathbb{E}_\theta f(Y_1) = \mu_\theta(f) + \Delta_n R(\Delta_n;\theta)$, (6.14) implies that

$$
\breve{a}_n(\theta)_0 = \mathbb{E}_\theta f(Y_1)\left(1 - \breve{a}_n(\theta)_1\right) = -\Delta_n K_f(\theta)\mu_\theta(f) + \Delta_n^{3/2} R(\Delta_n;\theta).
$$

□

Proof of Lemma 5.5. Define

$$
\begin{aligned}
g_1(\Delta_n, Y_i, Y_{i-1};\theta) &= f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1}), \qquad (6.15) \\
g_2(\Delta_n, Y_i, Y_{i-1};\theta) &= f(Y_{i-1})\left[f(Y_i) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(Y_{i-1})\right],
\end{aligned}
$$

and $H_n(\theta) = (n\Delta_n)^{-1} G_n(\theta) = (n\Delta_n)^{-1}\sum_{i=2}^n g(\Delta_n, Y_i, Y_{i-1};\theta)$, where $g = (g_1, g_2)^T$.

By the expansion (5.9)

$$
g_1(\Delta_n, Y_i, Y_{i-1};\theta) = f(Y_i) - f(Y_{i-1}) + \Delta_n K_f(\theta)\left[\mu_\theta(f) - f(Y_{i-1})\right] + \Delta_n^{3/2} R(\Delta_n, Y_{i-1};\theta), \quad (6.16)
$$

and, hence, by the law of large numbers for integrated diffusions (Lemma 4.1),

$$
\begin{aligned}
\frac{1}{n\Delta_n}\sum_{i=2}^n g_1(\Delta_n, Y_i, Y_{i-1};\theta) &= \frac{1}{n\Delta_n}\left[f(Y_n) - f(Y_1)\right] + \frac{1}{n}\sum_{i=2}^n K_f(\theta)\left[\mu_\theta(f) - f(Y_{i-1})\right] \\
&\quad + \Delta_n^{1/2}\frac{1}{n}\sum_{i=2}^n R(\Delta_n, Y_{i-1};\theta) \xrightarrow{\mathbb{P}_0} K_f(\theta)(\mu_\theta - \mu_0)(f).
\end{aligned}
$$

The second coordinate of $H_n(\theta)$ requires a considerably longer proof, because the contribution from the first term is not asymptotically negligible. To shorten the notation, we define $\mathcal{E}_i^n = \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta_0)\varepsilon_{1,i}$ and $\Xi_i^n = \partial_x f(X_{t_{i-1}^n}) b(X_{t_{i-1}^n};\theta_0)\xi_{1,i}$ and write the expansions (3.1) and (3.7) under the true probability measure $\mathbb{P}_0$ as

$$
\begin{aligned}
f(X_{t_i^n}) &= f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\mathcal{E}_i^n + \varepsilon_{2,i}, \qquad (6.17) \\
f(Y_i) &= f(X_{t_{i-1}^n}) + \Delta_n^{1/2}\Xi_i^n + \xi_{2,i}. \qquad (6.18)
\end{aligned}
$$

First note that by (6.18)

$$
g_2(\Delta_n, Y_i, Y_{i-1};\theta) = \left(f(X_{t_{i-2}^n}) + \Delta_n^{1/2}\Xi_{i-1}^n + \xi_{2,i-1}\right) g_1(\Delta_n, Y_i, Y_{i-1};\theta). \qquad (6.19)
$$

By inserting (6.18) into (6.15) and applying the expansion (5.9) of $\breve{a}_n(\theta)$, we find that

$$
\begin{aligned}
&g_1(\Delta_n, Y_i, Y_{i-1};\theta) = f(X_{t_{i-1}^n}) - f(X_{t_{i-2}^n}) \qquad (6.20) \\
&+ \ \Delta_n K_f(\theta)\left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] + \Delta_n^{1/2}\left(\Xi_i^n - \Xi_{i-1}^n\right) + \mathcal{R}_1(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta),
\end{aligned}
$$

where the remainder term $\mathcal{R}_1$ has the form

$$
\begin{aligned}
&\mathcal{R}_1(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta) \qquad (6.21) \\
&= \ (\xi_{2,i} - \xi_{2,i-1}) - \Delta_n^{3/2} K_f(\theta)\Xi_{i-1}^n - \Delta_n K_f(\theta)\xi_{2,i-1} + \Delta_n^{3/2} R(\Delta_n, Y_{i-1};\theta).
\end{aligned}
$$

Using (6.17) and inserting the definitions of $\varepsilon_{2,i}$ and $\xi_{2,i}$, (6.3) and (6.4), we obtain

$$f(X_{t_{i-1}^n}) - f(X_{t_{i-2}^n}) = \Delta_n^{1/2}\mathcal{E}_{i-1}^n + \varepsilon_{2,i-1} = \Delta_n \mathcal{L}_0 f(X_{t_{i-2}^n}) + A_{i-1}(\theta_0) + M_{i-1}(\theta_0), \quad (6.22)$$

where

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

Now, using (6.19),(6.20) and ((6.22)), we obtain the Δ-expansion

$$g_2(\Delta_n, Y_i, Y_{i-1};\theta) = \sum_{k=1}^{3} g_2^{(k)}(\Delta_n, Y_i, Y_{i-1};\theta),$$

where

$$\begin{aligned} & g_2^{(1)}(\Delta_n, Y_i, Y_{i-1};\theta) = f(X_{t_{i-2}^n}) \cdot g_1(\Delta_n, Y_i, Y_{i-1};\theta) \\ &= \Delta_n f(X_{t_{i-2}^n})\mathcal{L}_0 f(X_{t_{i-2}^n}) + f(X_{t_{i-2}^n}) M_{i-1}(\theta_0) \\ &\quad + \Delta_n K_f(\theta) f(X_{t_{i-2}^n})\left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] + \mathcal{R}_2^{(1)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta), \end{aligned}$$

$$\begin{aligned} g_2^{(2)}(\Delta_n, Y_i, Y_{i-1};\theta) &= \Delta_n^{1/2} \cdot \Xi_{i-1}^n \cdot g_1(\Delta_n, Y_i, Y_{i-1};\theta) \\ &= \Delta_n \left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right) \Xi_{i-1}^n + \mathcal{R}_2^{(2)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta) \end{aligned}$$

and

$$g_2^{(3)}(\Delta_n, Y_i, Y_{i-1};\theta) = \xi_{2,i-1} \cdot g_1(\Delta_n, Y_i, Y_{i-1};\theta) = \mathcal{R}_2^{(3)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta)$$

with

$$\begin{aligned} & \mathcal{R}_2^{(1)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta) = \\ & \quad f(X_{t_{i-2}^n}) A_{i-1}(\theta_0) + \Delta_n^{1/2} f(X_{t_{i-2}^n})\left(\Xi_i^n - \Xi_{i-1}^n\right) + f(X_{t_{i-2}^n}) \cdot \mathcal{R}_1(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta) \end{aligned}$$

and

$$\begin{aligned} & \mathcal{R}_2^{(2)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta) = \Delta_n^{1/2}\Xi_{i-1}^n \varepsilon_{2,i-1} \\ & + \Delta_n^{3/2}\Xi_{i-1}^n K_f(\theta)\left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] + \Delta_n \Xi_i^n \Xi_{i-1}^n + \Delta_n^{1/2}\Xi_{i-1}^n \mathcal{R}_1(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta). \end{aligned}$$

Collecting the terms,

$$\begin{aligned} & g_2(\Delta_n, Y_i, Y_{i-1};\theta) = \sum_{k=1}^{3} g_2^{(k)}(\Delta_n, Y_i, Y_{i-1};\theta) \quad (6.23) \\ &= \Delta_n f(X_{t_{i-2}^n})\mathcal{L}_0 f(X_{t_{i-2}^n}) + f(X_{t_{i-2}^n}) M_{i-1}(\theta_0) + \Delta_n K_f(\theta) f(X_{t_{i-2}^n})\left[\mu_\theta(f) - f(X_{t_{i-2}^n})\right] \\ &\quad + \Delta_n \left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right)\Xi_{i-1}^n + \mathcal{R}_2(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta), \end{aligned}$$

where the remainder term is

$$\mathcal{R}_2(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta) = \sum_{k=1}^{3} \mathcal{R}_2^{(k)}(\Delta_n, (X_s)_{s\in[t_{i-2}^n, t_i^n]}, \theta_0;\theta).$$

Now, tedious reasoning based on Lemma 9 in Genon-Catalot and Jacod (1993) shows that

$$\frac{1}{n\Delta_n}\sum_{i=2}^{n}\mathcal{R}_2(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=o_{\mathbb{P}_0}(1). \tag{6.24}$$

Under the additional rate assumption $n\Delta_n^2\to 0$, it can in a similar way be proved that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^{n}\mathcal{R}_2(\Delta_n,(X_s)_{s\in[t_{i-2}^n,t_i^n]},\theta_0;\theta)=o_{\mathbb{P}_0}(1). \tag{6.25}$$

The latter result is not needed in this proof, but it is necessary to show asymptotic normality in the proof of Theorem 5.6, so we state it here for convenience. To see that the strong rate assumption $n\Delta_n^2\to 0$ is necessary to obtain (6.25), we can, e.g., consider the last term in (6.21):

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^{n}\Delta_n^{3/2}R(\Delta_n,Y_{i-1};\theta)=\sqrt{n\Delta_n^2}\cdot\frac{1}{n}\sum_{i=2}^{n}R(\Delta_n,Y_{i-1};\theta).$$

As the proofs of (6.24) and (6.25) are both very long and not particularly enlightening, they are omitted.

To determine the limit in probability of the second coordinate of $H_n(\theta)$, we consider each term in (6.23) separately. By the ergodic theorem, see e.g. Lemma 3.1 in Jørgensen and Sørensen (2021),

$$\frac{1}{n}\sum_{i=2}^{n}f(X_{t_{i-2}^n})\mathcal{L}_0f(X_{t_{i-2}^n})\xrightarrow{\mathbb{P}_0}\mu_0(f\mathcal{L}_0f)$$

and

$$\frac{1}{n}\sum_{i=2}^{n}K_f(\theta)f(X_{t_{i-2}^n})\left[\mu_\theta(f)-f(X_{t_{i-2}^n})\right]\xrightarrow{\mathbb{P}_0}K_f(\theta)\left[\mu_0(f)\mu_\theta(f)-\mu_0(f^2)\right].$$

Furthermore, by definitions of $\mathcal{E}_i^n$ and Ξ_i^n,

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_0\left(\left(\mathcal{E}_i^n-\Xi_i^n\right)\Xi_i^n\;\middle|\;\mathcal{F}_{i-1}^n\right)=\frac{1}{n}\sum_{i=1}^{n}[\partial_xf(X_{t_{i-1}^n})b(X_{t_{i-1}^n};\theta_0)]^2\mathbb{E}_0\left((\varepsilon_{1,i}-\xi_{1,i})\xi_{1,i}\;\middle|\;\mathcal{F}_{i-1}^n\right),$$

and since $\xi_{1,i}\sim\mathcal{N}(0,1/3)$, (3.11) implies that

$$\mathbb{E}_0\left((\varepsilon_{1,i}-\xi_{1,i})\xi_{1,i}\;\middle|\;\mathcal{F}_{i-1}^n\right)=\mathbb{E}_0\left((\varepsilon_{1,i}-\xi_{1,i})\xi_{1,i}\right)=\frac{1}{6}, \tag{6.26}$$

so

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}_0\left(\left(\mathcal{E}_i^n-\Xi_i^n\right)\Xi_i^n\;\middle|\;\mathcal{F}_{i-1}^n\right)\xrightarrow{\mathbb{P}_0}\frac{1}{6}\mu_0\left([b(\,\cdot\,;\theta_0)\partial_xf]^2\right).$$

Finally, since by the definitions of $\varepsilon_{1,i}$ in (6.3) and of $\xi_{1,i}$ in (6.8) the difference $\varepsilon_{1,i}-\xi_{1,i}$ is Gaussian, Hölder's inequality and the ergodic theorem imply that

$$\begin{aligned}&\frac{1}{n^2}\sum_{i=1}^{n}\mathbb{E}_0\left((\mathcal{E}_i^n-\Xi_i^n)^2(\Xi_i^n)^2\;\middle|\;\mathcal{F}_{i-1}^n\right)\\ =\;&\frac{1}{n^2}\sum_{i=1}^{n}[\partial_xf(X_{t_{i-1}^n})b(X_{t_{i-1}^n};\theta_0)]^4\cdot\mathbb{E}_0\left((\varepsilon_{1,i}-\xi_{1,i})^2\,\xi_{1,i}^2\;\middle|\;\mathcal{F}_{i-1}^n\right)=o_{\mathbb{P}_0}(1).\end{aligned}$$

Therefore, by Lemma 9 in Genon-Catalot and Jacod (1993)

$$\frac{1}{n}\sum_{i=1}^{n}\left(\mathcal{E}_i^n-\Xi_i^n\right)\Xi_i^n \xrightarrow{\mathbb{P}_0} \frac{1}{6}\mu_0\left([b(\,\cdot\,;\theta_0)\partial_x f]^2\right).$$

By the similar arguments,

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

where we use that $\mathbb{E}_0\left(M_i(\theta_0)\,\middle|\,\mathcal{F}_{i-1}^n\right)=0$. Moreover, we use that, with $h(x)=\partial_x f(x)b(x;\theta_0)$, and using the conditional Itô isometry, Tonelli's theorem and (6.2),

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

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, we have verified (5.12).

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

$$H_n(\theta)=\frac{1}{n\Delta_n}\sum_{i=2}^{n} Z_{i-1}\left[f(Y_i)-Z_{i-1}^T\breve{a}_n(\theta)\right],$$

where $Z_{i-1}=(1,f(Y_{i-1}))^T$, which implies that

$$\partial_{\theta^T}H_n(\theta)=-\frac{1}{n\Delta_n}\sum_{i=2}^{n} Z_{i-1}Z_{i-1}^T\partial_{\theta^T}\breve{a}_n(\theta)=Z_n(f)A_n(\theta),$$

with $Z_n(f):=\frac{1}{n}\sum_{i=2}^n Z_{i-1}Z_{i-1}^T$ and $A_n(\theta):=-\Delta_n^{-1}\partial_{\theta^T}\breve{a}_n(\theta)$. By Lemma 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 expansion (5.9) of $\breve{a}_n(\theta)$, we see that

$$A_n(\theta)=\partial_{\theta^T}\begin{pmatrix}K_f(\theta)\mu_\theta(f)\\ -K_f(\theta)\end{pmatrix}+\Delta_n^{1/2}\partial_{\theta^T}R(\Delta_n;\theta)\to\begin{pmatrix}\partial_{\theta^T}[K_f(\theta)\mu_\theta(f)]\\ -\partial_{\theta^T}K_f(\theta)\end{pmatrix}=:A(\theta).$$

Hence, it follows that $\partial_{\theta^T}H_n(\theta)\xrightarrow{\mathbb{P}_0} Z(f)A(\theta)$. To argue that under Condition 5.4, the convergence is uniform over any compact subset $\mathcal{M}$ of Θ, note that

$$\begin{aligned}\sup_{\theta\in\mathcal{M}}\left\|\partial_{\theta^T}H_n(\theta)-Z(f)A(\theta)\right\|&\leq\sup_{\theta\in\mathcal{M}}\left\|Z_n(f)[A_n(\theta)-A(\theta)]\right\|+\sup_{\theta\in\mathcal{M}}\left\|[Z_n(f)-Z(f)]A(\theta)\right\|\\&\leq\left\|Z_n(f)\right\|\sup_{\theta\in\mathcal{M}}\left\|A_n(\theta)-A(\theta)\right\|+\left\|Z_n(f)-Z(f)\right\|\sup_{\theta\in\mathcal{M}}\left\|A(\theta)\right\|.\end{aligned}$$

Therefore, (5.13) follows because $\theta \mapsto A(\theta)$ and $\|\cdot\|$ are continuous, and because

$$\sup_{\theta\in\mathcal{M}} \|A_n(\theta) - A(\theta)\| = \Delta_n^{1/2} \sup_{\theta\in\mathcal{M}} \|\partial_{\theta^T} R(\Delta_n;\theta)\| \leq_{C(\mathcal{M})} \Delta_n^{1/2} \to 0.$$

□

Proof of Theorem 5.6. We use the notation introduced in the proof of Lemma 5.5. Because $\gamma(\theta_0,\theta_0) = 0$, the existence of a consistent sequence of G_n-estimators $(\hat\theta_n)$ follows from Lemma 5.5 and Theorem 2.5 in Jacod and Sørensen (2018). The eventual uniqueness in $\mathcal{K}$ follows from Lemma 5.5 and Theorem 2.7 in the same paper.

Asymptotic normality of $\hat\theta_n$ follows by a standard Taylor expansion argument (see e.g. Theorem 2.11 in Jacod and Sørensen (2018)) once we have established that

$$\sqrt{n\Delta_n}\cdot H_n(\theta_0) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2(0, V_0(f)). \tag{6.27}$$

From the expansion of $g_1(\Delta_n, Y_i, Y_{i-1};\theta)$ in (6.16), 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 the convergence in law holds by Lemma 4.2 because $f_1^* \in \mathscr{H}_0$.

Our proof that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n g_2(\Delta_n, Y_i, Y_{i-1};\theta_0) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mu_0\left([\partial_x U_0(f_2^*) + f\partial_x f]^2\, b^2(\,\cdot\,;\theta_0)\right)\right) \tag{6.28}$$

is based on the expansion of g_2 given by (6.23) and the observation that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n \Delta_n\left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right)\Xi_{i-1}^n = \frac{1}{6}\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n \Delta_n[\partial_x f(X_{t_{i-2}^n}) b(X_{t_{i-2}^n};\theta_0)]^2 + o_{\mathbb{P}_0}(1), \tag{6.29}$$

which follows from Lemma 9 in Genon-Catalot and Jacod (1993) using that $\left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right)\Xi_{i-1}^n = [\partial_x f(X_{t_{i-2}^n}) b(X_{t_{i-2}^n};\theta_0)]^2\,(\varepsilon_{1,i-1} - \xi_{1,i-1})\,\xi_{1,i-1}$ and that $\mathbb{E}_0\left(\left(\mathcal{E}_{i-1}^n - \Xi_{i-1}^n\right)\Xi_{i-1}^n \,\middle|\, \mathcal{F}_{i-2}^n\right) = \frac{1}{6}[\partial_x f(X_{t_{i-2}^n}) b(X_{t_{i-2}^n};\theta_0)]^2$, see (6.26).

Combining (6.29), (6.23) and the result (6.25) that the term involving the remainder term vanishes, we see that

$$\begin{aligned}
&\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n g_2(\Delta_n, Y_i, Y_{i-1};\theta_0) \\
&= \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right) + \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n f(X_{t_{i-1}^n}) M_i(\theta_0) + o_{\mathbb{P}_0}(1).
\end{aligned} \tag{6.30}$$

To gather the non-negligible terms in (6.30), we initially observe that

$$\frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n} f_2^*(X_s)\,\mathrm{d}s = \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right) \tag{6.31}$$
$$+ \quad \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n}\left[f_2^*(X_s) - f_2^*(X_{t_{i-1}^n})\right]\mathrm{d}s = \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right) + o_{\mathbb{P}_0}(1),$$

where we only use that $f_2^* \in \mathcal{C}_p^2(S)$. A proof that the second term in (6.31) is asymptotically negligible under $\mathbb{P}_0$ is contained in the proof of Proposition 3.4 in Jørgensen and Sørensen (2021). Furthermore, by Proposition 3.3 in the same paper, $\mathcal{L}_0\left(U_0(f_2^*)\right) = -f_2^*$ under Condition 5.4, and, therefore, by Itô's formula,

$$\begin{aligned} U_0(f_2^*)(X_t) &= U_0(f_2^*)(X_0) + \int_0^t \mathcal{L}_0(U_0(f_2^*))(X_s)\,\mathrm{d}s + \int_0^t \partial_x U_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s \\ &= U_0(f_2^*)(X_0) - \int_0^t f_2^*(X_s)\,\mathrm{d}s + \int_0^t \partial_x U_0(f_2^*)(X_s)b(X_s;\theta_0)dB_s. \end{aligned}$$

As a consequence,

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

and hence

$$\begin{aligned} &\frac{1}{\sqrt{n\Delta_n}}\sum_{i=2}^n g_2(\Delta_n, Y_i, Y_{i-1};\theta_0) \\ &= \sqrt{n\Delta_n}\left(\frac{1}{n}\sum_{i=1}^n f_2^*(X_{t_{i-1}^n})\right) + \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n f(X_{t_{i-1}^n})M_i(\theta_0) + o_{\mathbb{P}_0}(1) \\ &= \frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n}\left[\partial_x U_0(f_2^*)(X_s) + f(X_{t_{i-1}^n})\partial_x f(X_s)\right] b(X_s;\theta_0)dB_s + o_{\mathbb{P}_0}(1). \end{aligned}$$

At this point, the asymptotic normality in (6.28) can be shown by applying the central limit theorem for martingale difference arrays, see e.g. Hall and Heyde (1980) or Häusler and Luschgy (2015); for details see pp. 507-508 in Jørgensen and Sørensen (2021). The joint normality in (6.27) follows by the Cramér-Wold device. □

7 Concluding remarks and extensions

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 convergence of estimators of a parameter in the diffusion coefficient is $1/\sqrt{n}$. The optimal rates for integrated diffusion models under the high-frequency/infinite horizon scenario considered in this paper are not known, but the minimum contrast estimators in Gloter (2006) attain a rate of $1/\sqrt{n\Delta_n}$ for parameters in the drift and $1/\sqrt{n}$ for diffusion parameters under this scenario (similar to the rate optimal estimators for discretely observed diffusions in Sørensen (2024)), so presumably these rates are optimal. However, as we do not distinguish between drift and diffusion parameters in this paper, the $1/\sqrt{n\Delta_n}$ rate of our parameters is all we could hope for.

An interesting extension would be to introduce a jump component in the dynamics of (X_t). Such an 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. A general test for the presence of volatility jumps using change-point theory was proposed by Bibinger *et al.* (2017). How and whether the same principle can be applied for parametric inference is an interesting topic for future research.

Acknowledgement

Emil S. Jørgensen gratefully acknowledges financial support during a research stay from the Stevanovich Center for Financial Mathematics, University of Chicago.

References

Aït-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica*, **70**(1), 223–262.

Aït-Sahalia, Y. and Jacod, J. (2014). *High-Frequency Financial Econometrics*. Princeton University Press.

Baltazar-Larios, F. and Sørensen, M. (2010). Maximum likelihood estimation for integrated diffusion processes. In C. Chiarella and A. Novikov, editors, *Contemporary Quantitative Finance: Essays in Honour of Eckhard Platen*, pages 407–423. Springer.

Barndorff-Nielsen, O. and Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society*, **64**(2), 253–280.

Beskos, A., Papaspiliopoulos, O., Roberts, G., and Fearnhead, P. (2006). Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes (with discussion). *Journal of the Royal Statistical Society*, **68**(3), 333–382.

Bibby, B. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**(1/2), 17–39.

Bibinger, M., Jirak, M., and Vetter, M. (2017). Nonparametric change-point analysis of volatility. *Annals of Statistics*, **45**(4), 1542–1578.

Bladt, M. and Sørensen, M. (2014). Simple simulation of diffusion bridges with application to likelihood inference for diffusions. *Bernoulli*, **20**, 645–675. See also the following corrigendum.

Bollerslev, T. and Zhou, H. (2002). Estimating stochastic volatility diffusion using conditional moments of integrated volatility. *Journal of Econometrics*, **109**, 33–65.

Comte, F., Genon-Catalot, V., and Rozenholc, Y. (2009). Nonparametric adaptive estimation for integrated diffusions. *Stochastic Processes and their Applications*, **119**(3), 811–834.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

Ditlevsen, S. and Sørensen, M. (2004). Inference for observations of integrated diffusion processes. *Scandinavian Journal of Statistics*, **31**, 417–429.

Drimus, G. (2012). Options on realized variance by transform methods: a non-affine stochastic volatility model. *Quantitative Finance*, **12**(11), 1679–1694.

Forman, J. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, **35**, 438–465.

García-Portugués, E. and Sørensen, M. (2025). A family of toriodal diffusions with exact likelihood inference. *Bioketrika*. To appear, arXiv:2401.04689.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Ann. Inst. Henri Poincaré*, **29**(1), 119–151.

Genon-Catalot, V., Jeantheau, T., and Larédo, C. (2000). Stochastic volatility models as hidden markov models and statistical applications. *Bernoulli*, **6**(6), 1051–1079.

Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: Probability and Statistics*, **4**, 205–227.

Gloter, A. (2006). Parameter estimation for a discretely observed integrated diffusion process. *Scandinavian Journal of Statistics*, **33**, 83–104.

Gloter, A. and Gobet, E. (2008). Lamn property for hidden processes: The case of integrated diffusions. *Ann. Inst. Henri Poincaré*, **44**(1), 104–128.

Hall, P. and Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications*. Academic Press, New York.

Hansen, L. and Scheinkman, J. (1995). Back to the future: Generating moment implications for continuous-time markov processes. *Econometrica*, **63**(4), 767–804.

Heston, S. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies*, **6**(2), 327–343.

Häusler, E. and Luschgy, H. (2015). *Stable Convergence and Stable Limit Theorems*. Springer.

Jacod, J. and Protter, P. (2012). *Discretization of Processes*. Springer-Verlag.

Jacod, J. and Sørensen, M. (2018). A review of asymptotic theory of estimating functions. *Statistical Inference for Stochastic processes*, **21**, 415–434.

Jørgensen, E. and Sørensen, M. (2021). Prediction-based estimation for diffusion models with high-frequency data. *Japanese Journal of Statistics and Data Science*, **4**(1), 483–511.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scandinavian Journal of Statistics*, **24**, 211–229.

Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process. *Scandinavian Journal of Statistics*, **27**, 65–82.

Li, J. and Xiu, D. (2016). Generalized method of integrated moments for high-frequency data. *Econometrica*, **84**(4), 1613–1633.

Mancini, C. (2009). Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics*, **36**, 270–296.

Nelson, D. (1990). Arch models as diffusion approximations. *Journal of Econometrics*, **45**, 7–38.

Pardoux, E. and Veretennikov, A. Y. (2001). On the poisson equation and diffusion approximation. i. *Annals of Probability*, **29**(3), 1061–1085.

Pilipovic, P., Samson, A., and Ditlevsen, S. (2024). Parameter estimation in non-linear multivariate stochastic differential equations based on splitting schemes. *Ann. Statist.*, **52**, 848–8867.

Roberts, G. and Stramer, O. (2001). On inference for partially observed nonlinear diffusion models using the metropolis–hastings algorithm. *Biometrika*, **88**(3), 603–621.

Shoji, I. and Ozaki, T. (1998). A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika*, **85**(1), 240–243.

Sørensen, M. (2000). Prediction-based estimating functions. *Econometrics Journal*, **3**, 123–147.

Sørensen, M. (2011). Prediction-based estimating functions: review and new developments. *Brazilian Journal of Probability and Statistics*, **25**(3), 362–391.

Sørensen, M. (2012). Estimating functions for diffusion-type processes. In M. Kessler, A. Lindner, and M. Sørensen, editors, *Statistical Methods for Stochastic Differential Equations*, pages 1–107. CRC Press.

Sørensen, M. (2024). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint, arXiv:2401.04689.

Todorov, V. (2009). Estimation of continuous-time stochastic volatility models with jumps using high-frequency data. *Journal of Econometrics*, **148**(2), 131–148.

van der Meulen, F. and Schauer, M. (2017). Bayesian estimation of discretely observed multi-dimensional diffusion processes using guided proposals. *Electronic Journal of Statistics*, **11**, 2358–2396.

Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *Journal of Multivariate Analysis*, **41**(2), 220–242.