Prediction-based estimation for diffusion models with high-frequency data

Emil S. Jørgensen[1 2] Michael Sørensen[1 3]

July 27, 2020

Abstract. This paper obtains asymptotic results for parametric inference using prediction-based estimating functions when the data are high frequency observations of a diffusion process with an infinite time horizon. Specifically, the data are observations of a diffusion process at n equidistant time points $\Delta_n i$, and the asymptotic scenario is $\Delta_n \to 0$ and $n\Delta_n \to \infty$. For a useful and tractable classes of prediction-based estimating functions, existence of a consistent estimator is proved under standard weak regularity conditions on the diffusion process and the estimating function. Asymptotic normality of the estimator is established under the additional rate condition $n\Delta_n^3 \to 0$. The prediction-based estimating functions are approximate martingale estimating functions to a smaller order than what has previously been studied, and new non-standard asymptotic theory is needed. A Monte Carlo method for calculating the asymptotic variance of the estimators is proposed.

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

1 Introduction

Diffusion processes are often used to model stochastic dynamical systems. An especially successful application area is finance. These processes are defined in continuous time, but for most applications the system is only observed at discrete time points, so statistical methods for discretely observed diffusion processes is a very active area of research. In particular, the availability of high-frequency data has generated considerable interest in the asymptotic behaviour of estimators and test statistics as the time between consecutive observations tends to zero.

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

$$dX_t = a(X_t;\theta)dt + b(X_t;\theta)dB_t, \tag{1}$$

where (B_t) is standard Brownian motion, and the known functions a and b depend on a statistical parameter $\theta \in \Theta \subseteq \mathbb{R}^d$ to be estimated. We suppose that (X_t) takes values in

[1]Deptartment of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark
[2]IKEA Group, Älmhultsgatan 2, 215 86 Malmö, Sweden
[3]Corresponding author, michael@math.ku.dk

an open interval $(l, r) \subseteq \mathbb{R}$ and has a invariant distribution μ_θ. Moreover, (X_t) is assumed to be stationary under the probability measure $\mathbb{P}_\theta$, i.e. $X_0 \sim \mu_\theta$. Let the data be a single discretisation

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

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

$$n \to \infty, \quad \Delta_n \to 0, \quad n \cdot \Delta_n \to \infty, \tag{2}$$

where the time horizon $T_n = n\Delta_n$ tends to infinity with the number of observations.

Estimators are defined and studied within the framework of the prediction-based estimating functions, proposed by Sørensen (2000, 2011) as a versatile estimation framework, not least for non-Markovian diffusion-type models. They generalize the martingale estimating functions introduced by Bibby and Sørensen (1995). We show that the estimating functions considered in this paper are not approximate martingale estimating functions as defined in Sørensen (2017). However, for a two-dimensional predictor space, they are approximate martingale estimating functions of a smaller order than what has previously been studied, namely of order Δ_n rather than Δ_n^k for $k \geq 2$. We can still prove existence of consistent estimators, and by applying non-standard limit theory we establish asymptotically normality under mild regularity conditions and the additional rate assumption $n\Delta_n^3 \to 0$.

Examples from finance and simulation studies as well as more details on the theory and implementation issues can be found in Jørgensen (2017).

Parametric estimation for discretely observed diffusion processes has been investigated in many papers in the econometrics and statistics literature. Since exact maximum likelihood estimation is untractable for most diffusion models used in practice, a wide range of alternative methods have been proposed and applied successfully. The Markov property of diffusions 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 Sørensen (1995), Kessler and Sørensen (1999), Kessler (2000), Sørensen (2012), Jakobsen and Sørensen (2017)), likelihood expansions (Dacunha-Castelle and Florens-Zmirou (1986), Aït-Sahalia (2002), Li (2013)), Markov-chain Monte Carlo (Elerian *et al.* (2001), Eraker (2001), Roberts and Stramer (2001)) and simulated likelihood (Beskos *et al.* (2006), Beskos *et al.* (2009), Bladt *et al.* (2016)).

There is also a well developed literature on nonparametric estimation of the drift and diffusion coefficients from discrete time data. The problem was studied by Aït-Sahalia (1996), Hansen *et al.* (1998), Hoffmann (1999a), Gobet *et al.* (2004) and Comte *et al.* (2007) under the assumption of strict stationarity. Estimation for nonstationary, recurrent diffusion processes was considered by Bandi and Phillips (2003). Estimation of the diffusion coefficient with high-frequency observations on a finite time horizon was investigated by Genon-Catalot *et al.* (1992), Florens-Zmirou (1993), Hoffmann (1999a,b), Jacod (2000) and Renò (2008). Fan (2005) gives an excellent survey of nonparametric estimation with an extensive list of references.

The structure of the paper is as follows. In Section 2 we present the general notation used in the paper, define a tractable class of prediction-based estimating functions, and formulate our general assumption on (X_t). Section 3 is devoted to limit theorems for functionals $V_n(f) = n^{-1}\sum_{i=1}^n f(X_{t_{i-1}^n})$ and, in particular, a central limit theorem for f belonging to a large class of functions. The variance of the gaussian limit law involves the potential of f, which is considered in some detail. Asymptotic results are provided in Section 4. In Section 5 we propose Monte Carlo methods for determining the asymptotic variances obtained in Section 4. All proofs are deferred to Appendix A, and Appendix B contains some auxiliary results needed in the proofs.

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.

2.1 Notation

Our general notation is as follows:

1. The parameter of interest $\theta \in \Theta \subseteq \mathbb{R}^d$ 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. The invariant distribution is denoted by μ_θ. For short, we write $\mu_\theta(f) = \int_S f(x)\mu_\theta(dx)$ for functions $f : S \to \mathbb{R}$, and we denote the canonical norm on $\mathscr{L}^2(\mu_\theta)$ defined by $\|f\|_2 = \mu_\theta(f^2)^{1/2}$.
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 $Y \leq C \cdot Z$, $\mathbb{P}_{\theta_0}$-almost surely. We sometimes use a similar notation for real functions.

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 $\theta \in \Theta$ there exists a constant $C_\theta > 0$ such that, $|f(x;\theta)| \leq C_\theta(1+|x|^{C_\theta})$ 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.

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

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

$$|R(\Delta, x; \theta)| \leq F(x; \theta), \tag{3}$$

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 Prediction-based estimating functions

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

$$G_n(\theta) = \sum_{i=q}^{n} \sum_{j=1}^{N} \pi_{i-1,j} \left[f_j(X_{t_i^n}) - \breve{\pi}_{i-1,j}(\theta) \right], \tag{4}$$

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

$$\mathcal{P}_{i-1,j} = \text{span} \left\{ 1, f_j\left(X_{t_{i-1}^n}\right), \ldots, f_j\left(X_{t_{i-q_j}^n}\right) \right\} \subset \mathscr{L}^2(\mu_\theta) \tag{5}$$

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. Sørensen (2000) for applications to discretized stochastic volatility models, and Ditlevsen and 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

$$\mathbb{E}_\theta \left(\pi \left[f_j(X_{t_i^n}) - \breve{\pi}_{i-1,j}(\theta) \right] \right) = 0, \tag{6}$$

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 probability measure $\mathbb{P}_\theta$. Note that (6) implies that the estimating function (4) is unbiased, i.e.

$$\mathbb{E}_\theta \left(G_n(\theta) \right) = 0.$$

By restricting ourselves to predictor spaces of the form (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

$$Z_{i-1,j} = \left(1, f_j\left(X_{t_{i-1}^n}\right), \ldots, f_j\left(X_{t_{i-q_j}^n}\right) \right)^T \tag{7}$$

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

$$\mathbb{E}_\theta \left[Z_{q_j-1,j} f_j\left(X_{t_{q_j}^n}\right) \right] - \mathbb{E}_\theta \left[Z_{q_j-1,j} Z_{q_j-1,j}^T \right] \breve{a}_n(\theta)_j = 0. \tag{8}$$

Note that in the simplest case where $q_j = 0$, $\mathcal{P}_{i-1,j} = \text{span}\{1\}$ and it follows immediately from the normal equations (6) that $\breve{\pi}_{i-1,j}(\theta) = \mu_\theta(f_j)$.

We obtain an estimator 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 $n \to \infty$.

Remark 2.1. If we define an equivalence relation $\sim$ on the set of estimating functions of the form (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 4.

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.

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

A stationary Markov process (X_t) is said to be *ρ-mixing* if $\rho_X(t) \to 0$ as $t \to 0$, where

$$\rho_X(t) = \rho\left(\sigma(X_0), \sigma(X_t)\right), \tag{9}$$

with ρ denoting correlation. A review of mixing properties for stationary Markov processes can be found in Genon-Catalot *et al.* (2000). Here easily checked conditions for ρ-mixing of one-dimensional diffusion processes are given.

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

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

Obviously, $P_t^\theta : \mathscr{L}^2(\mu_\theta) \to \mathscr{L}^2(\mu_\theta)$, 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 by

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

whenever the limit $\mathcal{A}_\theta f$ exists in $\mathscr{L}^2(\mu_\theta)$. Let $\mathcal{D}_{\mathcal{A}_\theta}$ denote the *domain* of $\mathcal{A}_\theta$. For a weak solution of (1) satisfying Condition 2.2 below, $\mathcal{C}_p^2(S) \subseteq \mathcal{D}_{\mathcal{A}_\theta}$, and for all $f \in \mathcal{C}_p^2(S)$ it holds that $\mathcal{A}_\theta f = \mathcal{L}_\theta f$, where

$$\mathcal{L}_\theta f(x) = a(x;\theta)\partial_x f(x) + \frac{1}{2} b^2(x;\theta)\partial_x^2 f(x); \tag{10}$$

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 *et al.* (2000).

2.4 Assumptions

To derive asymptotic results for diffusion models of the general form (1), we impose some mild dependence and regularity conditions on (X_t).

Condition 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 and 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* $k \geq 1$.

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.

The following condition on the true parameter value θ_0 is essential to the asymptotic theory. Here $\mathrm{int}(\Theta)$ denotes the interior of Θ.

Condition 2.3. *The parameter $\theta \in \Theta \subset \mathbb{R}^d$, and it holds that $\theta_0 \in \mathrm{int}(\Theta)$.*

3 Limit theory for discretized diffusions

This section is devoted to limit theorems for functionals of the form

$$V_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(X_{t_{i-1}^n}), \tag{11}$$

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.

First we state the law of large numbers, which follows from the continuous-time ergodic theorem; see Corollary 10.9 in Kallenberg (2002).

Lemma 3.1. *Let $f \in \mathcal{C}_p^1(S)$. Then,*

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

A central limit theorem requires stronger regularity assumptions on f. In the following we define a suitable class of functions for our purpose. Our only application of the central limit theorem in this paper is to establish asymptotic normality of G_n-estimators in Section 4. Since $\mathbb{E}_\theta\left(G_n(\theta)\right) = 0$, we can restrict attention to functions $f : S \to \mathbb{R}$ for which $\mu_\theta(f) = 0$ for the remainder of this section.

The variance of the Gaussian limit distribution in the central limit theorem involves the potential of the function f with $\mu_\theta(f) = 0$. The *potential operator* is defined by

$$U_\theta(f)(x) = \int_0^\infty P_t^\theta f(x)\,\mathrm{d}t. \tag{12}$$

To identify a (partial) domain for the operator $f \mapsto U_\theta(f)$, we use that the generator $\mathcal{A}_\theta$ of (X_t) has a spectral gap $\lambda > 0$ under Condition 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 3.2. *Let $f \in \mathscr{L}_0^2(\mu_\theta)$. Then under Condition 2.2*

$$\left\| P_t^\theta f \right\|_2 \leq e^{-\lambda t} \left\| f \right\|_2 \tag{13}$$

for all $t \geq 0$.

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

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

is well-defined. It is obviously linear. 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).

For the central limit theorem, we restrict ourselves to the set of functions

$$\mathscr{H}_\theta^2 = \left\{f \in \mathcal{C}_p^2(S) : \mu_\theta(f) = 0, U_\theta(f) \in \mathcal{C}_p^2(S)\right\}, \tag{14}$$

which ensures that $\mathcal{A}_\theta(U_\theta(f)) = \mathcal{L}_\theta(U_\theta(f))$ 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 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 $\mathcal{L}_\theta$ is the generator of (X_t) given by the differential operator (10)*. Moreover,*

$$\|U_\theta(f)\|_2 \leq \lambda^{-1} \|f\|_2. \tag{15}$$

With Proposition 3.3 in place, we obtain the following central limit theorem. 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 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

$$\mathcal{V}_0(f) = \mu_0 \left([\partial_x U_0(f) b(\,\cdot\,; \theta_0)]^2 \right) = 2\mu_0 \left(f U_0(f) \right). \tag{16}$$

Remark 3.5. Compared to the low-frequency sampling scenario where $\Delta_n = \Delta > 0$, the integral defining $U_\theta(f)$ in (12) 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 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).

4 Asymptotic theory

In this section we present our main asymptotic results for prediction-based estimators. The main proof is based on general asymptotic theory for estimating functions in Jacod and Sørensen (2018); see also Sørensen (2012). For the most part, we restrict the discussion to estimating functions of the form (4) with $N = 1$ and, for simplicity, write

$$G_n(\theta) = \sum_{i=q}^{n} \pi_{i-1} \left[f(X_{t_i^n}) - \breve{\pi}_{i-1}(\theta) \right], \tag{17}$$

$\mathcal{P}_{i-1}$ for the corresponding predictor spaces and so on for objects in Section 2.2 that depend on j. The extension to multiple predictor functions $\{f_j\}_{j=1}^N$ is considered in Section 4.3.

4.1 Simple predictor spaces

The simplest class of estimating functions of the form (17) is obtained for $q = 0$, in which case $\mathcal{P}_{i-1} = \operatorname{span}\{1\}$. The orthogonal projection is $\breve{\pi}_{i-1}(\theta) = \mu_\theta(f)$, and the one-dimensional predictor space $\mathcal{P}_{i-1}$ enables us to estimate one real parameter $\theta \in \Theta \subseteq \mathbb{R}$. Therefore, we study the one-dimensional estimating function

$$G_n(\theta) = \sum_{i=1}^{n} \left[f(X_{t_i^n}) - \mu_\theta(f) \right]. \tag{18}$$

Such estimating functions were studied by Kessler (2000).

We easily identify conditions that ensure consistency and asymptotic normality of G_n-estimators

Condition 4.1. *Suppose that*

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

Theorem 4.2. *Assume Condition 4.1 and the identifiability condition $\partial_\theta \mu_\theta(f) \neq 0$ for all $\theta \in \Theta$. Define $\kappa(\theta) = \mu_\theta(f)$. Then the following assertions hold.*

- *There exists a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ which, as $n \to \infty$, is uniquely given by $\hat{\theta}_n = \kappa^{-1}\left(\frac{1}{n}\sum_{i=1}^n f(X_{t_i^n})\right)$ with $\mathbb{P}_0$-probability approaching one.*
- *If, moreover, $n\Delta_n^3 \to 0$, then*
$$\sqrt{n\Delta_n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \left[\partial_\theta \mu_{\theta_0}(f)\right]^{-2} \mathcal{V}_0(f)\right), \tag{19}$$
where $\mathcal{V}_0(f) = 2\mu_0(f^* U_0(f^*))$.

4.2 1-lag predictor spaces

The inclusion of past observations into the predictor space $\mathcal{P}_{i-1}$ raises the mathematical complexity dramatically. We show 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 it follows from the normal equations (8) that
$$\breve{\pi}_{i-1}(\theta) = \breve{a}_n(\theta)_0 + \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}),$$
where $\breve{a}_n(\theta)_0$ and $\breve{a}_n(\theta)_1$ are uniquely determined by the moment conditions
$$\begin{aligned} \breve{a}_n(\theta)_0 &= \mu_\theta(f)\left(1 - \breve{a}_n(\theta)_1\right), \\ \breve{a}_n(\theta)_1 &= \frac{\mathbb{E}_\theta\left[f(X_0)f(X_{\Delta_n})\right] - \left[\mu_\theta(f)\right]^2}{\mathbb{V}\mathrm{ar}_\theta f(X_0)}, \end{aligned}$$
and consistent with a two-dimensional predictor space $\mathcal{P}_{i-1}$, we suppose that $d = 2$ and study the estimating function
$$G_n(\theta) = \sum_{i=1}^n \begin{pmatrix} 1 \\ f(X_{t_{i-1}^n}) \end{pmatrix} \left[f(X_{t_i^n}) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n})\right]. \tag{20}$$

As part of the proof of Lemma 4.4 below, we show that the projection coefficient $\breve{a}_n(\theta)$ has an expansion
$$\breve{a}_n(\theta) = \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \Delta_n \begin{pmatrix} -K_f(\theta)\mu_\theta(f) \\ K_f(\theta) \end{pmatrix} + \Delta_n^2 R(\Delta_n;\theta), \tag{21}$$
where $|R(\Delta_n;\theta)| \leq C(\theta)$ and
$$K_f(\theta) = \frac{\mu_\theta(f\mathcal{L}_\theta f)}{\mathbb{V}\mathrm{ar}_\theta f(X_0)}. \tag{22}$$
This observation enables us to formulate a set of regularity conditions on G_n for the asymptotic theory:

Condition 4.3. *Suppose that*

- $f \in \mathcal{C}_p^4(S)$,
- $f_1^*(x) = K_f(\theta_0)\left[\mu_0(f) - f(x)\right] \in \mathscr{H}_0^2$,
- $f_2^*(x) = f(x)\left[\mathcal{L}_0 f(x) - f_1^*(x)\right] \in \mathscr{H}_0^2$,
- $(\theta \mapsto \mu_\theta(f)) \in \mathcal{C}^1$, $(\theta \mapsto K_f(\theta)) \in \mathcal{C}^1$ *and in* (21)

$$\sup_{\theta \in \mathcal{M}} \|\partial_{\theta^T} R(\Delta_n;\theta)\| \leq C(\mathcal{M}), \tag{23}$$

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

The matrix norm $\|\cdot\|$ in (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 4.5. As the proof is somewhat long, we formulate it as a separate result.

Lemma 4.4. *Assume that Condition 4.3 holds. Then, for any* $\theta \in \Theta$,

$$(n\Delta_n)^{-1} G_n(\theta) \xrightarrow{\mathbb{P}_0} \gamma(\theta_0;\theta) = \begin{pmatrix} K_f(\theta)(\mu_\theta - \mu_0)(f) \\ \mu_0(f\mathcal{L}_0 f) - K_f(\theta)\left[\mu_0(f^2) - \mu_0(f)\mu_\theta(f)\right] \end{pmatrix} \tag{24}$$

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

$$\sup_{\theta \in \mathcal{M}} \left\|(n\Delta_n)^{-1}\partial_{\theta^T} G_n(\theta) - W(\theta)\right\| \xrightarrow{\mathbb{P}_0} 0 \tag{25}$$

where

$$W(\theta) = \begin{pmatrix} 1 & \mu_0(f) \\ \mu_0(f) & \mu_0(f^2) \end{pmatrix} \begin{pmatrix} \partial_{\theta_1}\left[K_f(\theta)\mu_\theta(f)\right] & \partial_{\theta_2}\left[K_f(\theta)\mu_\theta(f)\right] \\ -\partial_{\theta_1} K_f(\theta) & -\partial_{\theta_2} K_f(\theta) \end{pmatrix}. \tag{26}$$

Theorem 4.5. *Assume Condition 4.3 and suppose that* $W(\theta_0)$ *is non-singular and that the following identifiability condition is satisfied*

$$\gamma(\theta_0;\theta) \neq 0 \quad \textit{for all } \theta \neq \theta_0.$$

Then the following assertions hold:

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

$$\sqrt{n\Delta_n}\left(\hat{\theta}_n - \theta_0\right) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2\left(0, \left[W(\theta_0)^{-1}\mathcal{V}_0(f)(W(\theta_0)^{-1})^T\right]\right), \tag{27}$$

wheren $W(\theta_0))$ is given by (26) and

$$
\begin{aligned}
\mathcal{V}_0(f)_{11} &= \mu_0\left([\partial_x U_0(f_1^*)b(\,\cdot\,;\theta_0)]^2\right) = 2\mu_0\left(f_1^* U_0(f_1^*)\right) \\
\mathcal{V}_0(f)_{12} &= \mathcal{V}_0(f)_{21} = \mu_0\left(\partial_x U_0(f_1^*)\left[\partial_x U_0(f_2^*) + ff'\right] b^2(\,\cdot\,;\theta_0)\right) \\
&= \mu_0\left(f_1^* U_0(f_2^*) + f_2^* U_0(f_1^*)\right) + \mu_0\left(\partial_x U_0(f_1^*) ff' b^2(\,\cdot\,;\theta_0)\right) \\
\mathcal{V}_0(f)_{22} &= \mu_0\left(\left[\partial_x U_0(f_2^*) + ff'\right]^2 b^2(\,\cdot\,;\theta_0)\right) \\
&= 2\mu_0\left(f_2^* U_0(f_2^*)\right) + \mu_0\left([ff'b(\,\cdot\,;\theta_0)]^2\right) + 2\mu_0\left(\partial_x U_0(f_2^*) ff' b^2(\,\cdot\,;\theta_0)\right).
\end{aligned}
$$

Remark 4.6. If we denote the estimating function (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 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 Sørensen (2017). In particular, the proof of asymptotic normality in Theorem 4.5 requires extra work, because the remainder term obtained by compensating G_n is non-negligible.

4.3 Multiple predictor functions and optimal estimation

Estimating functions with multiple predictor functions,

$$G_n(\theta) = \sum_{i=q}^{n}\sum_{j=1}^{N} \pi_{i-1,j}\left[f_j(X_{t_i^n}) - \breve{\pi}_{i-1,j}(\theta)\right] \tag{28}$$

appear frequently in practice. In the following, we indicate how to extend the asymptotic theory from estimating functions with a single predictor function (17) to the more general case (28) and briefly consider optimal estimation in relation to over-identification of the parameter $\theta \in \Theta \subset \mathbb{R}^d$.

To extend the proof in Appendix A from estimating functions with a single predictor function (17) to estimating functions of the more general form (28), we consider the more compact vector representation of the estimating functions (28)

$$G_n(\theta) = A_n(\theta)\sum_{i=q}^{n} Z_{i-1}\left[F(X_{t_i^n}) - \breve{\Pi}_{i-1}(\theta)\right], \tag{29}$$

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

$$Z_{i-1} = \begin{pmatrix} Z_{i-1,1} & 0_{q_1+1} & \cdots & 0_{q_1+1} \\ 0_{q_2+1} & Z_{i-1,2} & \cdots & 0_{q_2+1} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{q_N+1} & 0_{q_N+1} & \cdots & Z_{i-1,N} \end{pmatrix}. \tag{30}$$

Recall that $Z_{i-1,j}$ denotes the column vector (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 (30) 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.

To prove asymptotic results for the more general estimating equations, impose the condition that $A_n(\theta) \to A(\theta)$ as $n \to \infty$ and examine, by methods analogous to those used above, 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 $v_{n,k_j}^{(j)} \to 0$ at appropriate rates, e.g. $v_{n,k_j}^{(j)} = n^{-1}$ or $v_{n,k_j}^{(j)} = (n\Delta_n)^{-1}$.

The 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 $d < \bar{d}$, see Godambe and Heyde (1987) and Sørensen (2011).

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 propose a Monte Carlo-based method for calculating the difficult parts of the asymptotic variance (or covariance matrix) for the estimators derived in Sections 4.1 and 4.2. Moreover, we derive an upper bound for $\mathrm{AVAR}(\hat{\theta}_n)$ for estimating functions (18) and show that it is exact when estimating the mean of an Ornstein-Uhlenbeck process.

Terms in the asymptotic variance that are integrals of known functions with respect to the invariant measure can be found by standard methods. The difficult parts of the asymptotic variance are integrals with respect to the invariant measure that involve the potential. In the expression for $\mathcal{V}_0(f)$ in Theorem 4.5 there are terms of the form $\mu_\theta(f_1 \partial_x U_\theta(f_2))$, where $f_2 \in \mathscr{H}_\theta^2$. If we assume that the invariant measure μ_θ has a density ν_θ with respect to Lebesgue measure on the state space $S = (\ell, r)$ $(-\infty \leq \ell < r \leq \infty)$, which holds under weak regularity conditions, then it follows by integration by parts that

$$\begin{aligned}\mu_\theta\left(f_1\partial_x U_\theta(f_2)\right) &= \int_\ell^r f_1(x)\nu_\theta(x)\partial_x U_\theta(f_2)(x)dx \\ &= \nu_\theta(r)f_1(r)U_0(f_2)(r) - \nu_\theta(\ell)f_1(\ell)U_0(f_2)(\ell) - \mu_\theta\left(U_0(f_2)\left[f_1' + f_1(\log\nu_\theta)'\right]\right),\end{aligned}$$

where the function values at the end-points may have to be interpreted as limits and will often be equal to zero. Now an inspection of the expressions for the asymptotic variance in Theorems 4.2 and 4.5 shows that all difficult terms are of the form $\mu_\theta(g_1 U_\theta(g_2))$ with $g_1 \in \mathscr{L}^2(\mu_\theta)$ and $g_2 \in \mathscr{H}_\theta^2$, and in the following we propose a Monte Carlo method for calculating such terms.

For the construction we suppose that $\{T_i\}$ is a sequence of independent random variables defined on an auxiliary probability space $(\Omega', \mathcal{F}', \mathbb{P}'_\gamma)$ such that $T_i \sim \exp(\gamma)$ and consider the product extension

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

Obviously, $\tilde{\mathbb{E}}_{\theta,\gamma} f(X_0)) = \mathbb{E}_\theta f(X_0)$ for any $f \in \mathscr{L}^1(\mu_\theta)$. For any $g_1 \in \mathscr{L}^2(\mu_\theta)$ and $g_2 \in \mathscr{H}_\theta^2$,

$$\begin{aligned}\mu_\theta\left(g_1 U_\theta(g_2)\right) &= \int_S g_1(x)\left(\int_0^\infty P_t^\theta g_2(x)\,\mathrm{d}t\right)\mu_\theta(dx) \\ &= \int_0^\infty\left(\int_S g_1(x)P_t^\theta g_2(x)\mu_\theta(dx)\right)\mathrm{d}t \\ &= \int_0^\infty\left(\int_S \mathbb{E}_\theta\left(g_1(X_0)g_2(X_t) \mid X_0 = x\right)\mu_\theta(dx)\right)\mathrm{d}t \\ &= \int_0^\infty \mathbb{E}_\theta\left(g_1(X_0)g_2(X_t)\right)\mathrm{d}t \\ &= \gamma^{-1}\int_0^\infty e^{\gamma t}\tilde{\mathbb{E}}_{\theta,\gamma}\left(g_1(X_0)g_2(X_t)\right)\gamma e^{-\gamma t}\,\mathrm{d}t \\ &= \gamma^{-1}\int_0^\infty \tilde{\mathbb{E}}_{\theta,\gamma}\left(e^{\gamma T_i}g_1(X_0)g_2(X_{T_i}) \,\middle|\, T_i = t\right)\gamma e^{-\gamma t}\,\mathrm{d}t \\ &= \gamma^{-1}\tilde{\mathbb{E}}_{\theta,\gamma}\left[e^{\gamma T_i}g_1(X_0)g_2(X_{T_i})\right],\end{aligned}$$

where we have used Fubini's theorem and the fact that (X_t) and T_i are independent on $\tilde{\Omega}$ under $\tilde{\mathbb{P}}_{\theta,\gamma}$.

As a consequence, if $(X_t^{(i)})$ are independent trajectories of (X_t) under $\mathbb{P}_\theta$, the estimator

$$\gamma^{-1}\frac{1}{K}\sum_{i=1}^K e^{\gamma T_i} g_1\left(X_0^{(i)}\right) g_2\left(X_{T_i}^{(i)}\right) \tag{31}$$

converges $\tilde{\mathbb{P}}_{\theta,\gamma}$-almost surely to $\mu_\theta\left(g_1 U_\theta(g_2)\right)$ as $K \to \infty$ for any $g_1 \in \mathscr{L}^2(\mu_\theta)$ and $g_2 \in \mathscr{H}_\theta^2$.

5.1 Simple predictor spaces

Let us consider the estimator in Section 4.1 in detail. By Theorem 4.2,

$$\mathrm{AVAR}(\hat{\theta}_n) = \frac{2\mu_0\left(f^* U_0(f^*)\right)}{[\partial_\theta \mu_0(f)]^2} \tag{32}$$

with $f^* = f - \mu_0(f)$. Thus the following algorithm can be used to estimate the asymptotic variance:

MONTE CARLO ESTIMATION OF $\mathrm{AVAR}(\hat{\theta}_n)$

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

$$\widehat{\mathrm{AVAR}}(\hat{\theta}_n) = 2 \cdot [\partial_\theta \mu_{\hat{\theta}_n}(f)]^{-2} \gamma^{-1} \frac{1}{K} \sum_{i=1}^{K} e^{\gamma T_i} \hat{f}^* \left(X_0^{(i)} \right) \hat{f}^* \left(X_{T_i}^{(i)} \right), \tag{33}$$

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

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

Proposition 5.1. *Suppose that* (X_t) *and* $G_n(\theta)$ *satisfy 2.2 and 4.1, respectively, and let* λ_0 *denote the spectral gap of* (X_t) *under* $\mathbb{P}_0$. *Then,*

$$\mathrm{AVAR}(\hat{\theta}_n) \leq \frac{2\, \mathbb{V}\mathrm{ar}_0 f(X_0)}{\lambda_0 \, [\partial_\theta \mu_{\theta_0}(f)]^2}. \tag{34}$$

Example 5.2. The Ornstein-Uhlenbeck process

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

with $\kappa, \xi > 0$ and $\eta \in \mathbb{R}$, satisfies Condition 2.2. The invariant distribution is $\mathcal{N}\left(\eta, \frac{\xi^2}{2\kappa}\right)$.

Estimation of η (with κ and ξ are known) provides an illustrative example where the upper bound of in (34) is attained. We choose $f(x) = x$, and by direct calculation,

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

As a consequence,

$$\mu_0 \left(f^* U_0(f^*)\right) = \frac{1}{\kappa} \int_{\mathbb{R}} (x - \eta_0)^2 \mu_0(dx) = \frac{\xi^2}{2\kappa^2} \tag{35}$$

and

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

The bound (34) is attained because $\mathbb{V}\mathrm{ar}_0(X_0) = \frac{\xi^2}{2\kappa}$ and $\lambda_0 = \kappa$.

References

Aït-Sahalia, Y. (1996). Nonparametric pricing of interest rate derivative securities. *Econometrica*, **64**(3), 527–560.

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

Bandi, F. and Phillips, P. (2003). Fully nonparametric estimation of scalar diffusion models. *Econometrica*, **71**(1), 241–283.

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.

Beskos, A., Papaspiliopoulos, O., and Roberts, G. (2009). Monte carlo maximum likelihood estimation for discretely observed diffusion processes. *Annals of Statistics*, **37**(1), 223–245.

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

Bladt, M., Finch, S., and Sørensen, M. (2016). Simulation of multivariate diffusion bridges. *J.R. Statist. Soc. B*, **78**, 343–369.

Comte, F., Genon-Catalot, V., and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **13**(2), 514–543.

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.

Doukhan, P. (1994). *Mixing: Properties and Examples*. Lecture Notes in Statistics 85. Springer-Verlag.

Elerian, O., Chib, S., and Shephard, N. (2001). Likelihood inference for discretely observed nonlinear diffusions. *Econometrica*, **69**(4), 959–993.

Eraker, B. (2001). Mcmc analysis of diffusion models with application to finance. *Journal of Business & Economic Statistics*, **19**(2), 177–191.

Fan, J. (2005). A selective overview of nonparametric methods in financial econometrics. *Statistical Science*, **20**(4), 317–337.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, **20**, 547–557.

Florens-Zmirou, D. (1993). On estimating the diffusion coefficient from discrete observations. *Journal of Applied Probability*, **30**(4), 790–804.

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., Larédo, C., and Picard, D. (1992). Non-parametric estimation of the diffusion coefficient by wavelet methods. *Scandinavian Journal of Statistics*, **19**(4), 317–335.

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.

Gobet, E., Hoffmann, M., and Reiß, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Annals of Statistics*, **32**(5), 2223–2253.

Godambe, V. and Heyde, C. (1987). Quasi-likelihood and optimal estimation. *International Statistical Review*, **55**(3), 231–244.

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.

Hansen, L., Scheinkman, J., and Touzi, N. (1998). Spectral methods for identifying scalar diffusions. *Journal of Econometrics*, **86**(1), 1–32.

Hoffmann, M. (1999a). Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**(1), 135–163.

Hoffmann, M. (1999b). l_p estimation of the diffusion coefficient. *Bernoulli*, **5**(3), 447–481.

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

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics*, **27**, 83–96.

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

Jakobsen, N. and Sørensen, M. (2017). Efficient estimation for diffusions sampled at high frequency over a fixed time interval. *Bernoulli*, **23**(3), 1874–1910.

Jørgensen, E. (2017). *Diffusion Models Observed at High Frequency and Applications in Finance*. Ph.D. thesis, Department of Mathematical Sciences, University of Copenhagen.

Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer-Verlag.

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.

Kessler, M. and Sørensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, **5**(2), 299–314.

Li, C. (2013). Maximum-likelihood estimation for diffusion processes via closed-form density expansions. *Annals of Statistics*, **41**(3), 1350–1380.

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

Renò, R. (2008). Nonparametric estimation of the diffusion coefficient of stochastic volatility models. *Econometric Theory*, **24**(5), 1174–1206.

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

Rudin, W. (1987). *Real and Complex Analysis*. McGraw-Hill.

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. (2017). Efficient estimation for ergodic diffusions sampled at high frequency. Working paper.

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

Appendix A: Proofs

Proof of Lemma 3.2. The diffusion process (X_t) is reversible under Condition 2.2, so by Theorem 2.4 and Theorem 2.6 in Genon-Catalot *et al.* (2000) $\left\|P_t^\theta f\right\|_2 \leq \rho_X(t) \left\|f\right\|_2 = e^{-\lambda t} \left\|f\right\|_2$, for any $f \in \mathscr{L}_0^2(\mu_\theta)$, where $\lambda > 0$ denotes the spectral gap of $\mathcal{A}_\theta$. □

Proof of Proposition 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 all $n \in \mathbb{N}$ 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. $\|\cdot\|_2$. The latter equality holds because $\left\|P_n^\theta f\right\|_2 \leq \|f\|_2\, e^{-\lambda n} \to 0$.

By Jensen's inequality, Fubini's theorem and Lemma, 3.2 $U_\theta^{(n)}(f)$ converges to $U_\theta(f)$ in $\mathscr{L}^2(\mu_\theta)$ as $n \to \infty$:

$$\begin{aligned}
\left\|U_\theta(f) - U_\theta^{(n)}(f)\right\|_2^2 &= \int_S \left(\int_0^\infty 1\{t \geq n\}\lambda^{-1} e^{\lambda t} P_t^\theta f(x) \lambda e^{-\lambda t}\,\mathrm{d}t\right)^2 \mu_\theta(dx) \\
&\leq \int_S \left(\int_0^\infty 1\{t \geq n\}\lambda^{-2} e^{2\lambda t}\left(P_t^\theta f(x)\right)^2 \lambda e^{-\lambda t}\,\mathrm{d}t\right) \mu_\theta(dx) \\
&= \lambda^{-1}\int_n^\infty e^{\lambda t}\cdot ||P_t^\theta f||_2^2\,\mathrm{d}t \\
&\leq \lambda^{-1}\|f\|_2^2 \int_n^\infty e^{-\lambda t}\,\mathrm{d}t = \lambda^{-2}\|f\|_2^2\, e^{-\lambda n} \to 0.
\end{aligned}$$

Taking $n = 0$, we obtain (15). 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). □

Proof of Proposition 3.4. The proof is an application of the central limit theorem for martingales. For completeness and because we need to extend the result in a non-standard way later, we give the proof. First, note that

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

where we will show that

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^n \int_{(i-1)\Delta_n}^{i\Delta_n}\left[f(X_s) - f(X_{t_{i-1}^n})\right]\mathrm{d}s = o_{\mathbb{P}_0}(1). \tag{36}$$

With $A_i := \int_{(i-1)\Delta_n}^{i\Delta_n}\left[f(X_s) - f(X_{t_{i-1}^n})\right]\mathrm{d}s$, Fubini's theorem combined with B.2 implies that

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

for a generic function $F(x;\theta_0)$ of polynomial growth in x. Since $n\Delta_n^3 \to 0$, it follows by Lemma 3.1 that

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

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 B.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 conclusion (36) now follows from Lemma 9 in Genon-Catalot and Jacod (1993).

To apply the central limit theorem for martingales, note that Proposition 3.3 and Itô's formula applied to $U_0(f)$ imply that

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

so

$$\frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n} f(X_s)\,\mathrm{d}s = \frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n} \partial_x U_0(f)(X_s)b(X_s;\theta_0)dB_s + o_{\mathbb{P}_0}(1). \tag{37}$$

The stochastic integral is a true martingale under $\mathbb{P}_0$ and by the ergodic theorem

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

In conclusion,

$$\begin{aligned}\sqrt{n\Delta_n}V_n(f) &= \frac{1}{\sqrt{n\Delta_n}}\int_0^{n\Delta_n} \partial_x U_0(f)(X_s)b(X_s;\theta_0)dB_s + o_{\mathbb{P}_0}(1) \qquad (38)\\ &\xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mu_0\left([\partial_x U_0(f)b(\,\cdot\,;\theta_0)]^2\right)\right),\end{aligned}$$

where convergence in law under $\mathbb{P}_0$ follows from the continuous-time martingale central limit theorem (e.g. Theorem 6.31 in Häusler and Luschgy (2015)) or the central limit theorem for martingale arrays (e.g. Theorem 3.2 in Hall and Heyde (1980)). The conditional Lyapunov condition can be verified as in the proof of Theorem 4.5.

The alternative expression for the asymptotic variance $\mathcal{V}_0(f)$ in (16) follows because with $g(x) = U_0(f)$ and $b_0(x) = b(x;\theta_0)$ it follows from Proposition 3.3 that

$$2\mu_0(fg) = -\mu_0\left(\mathcal{L}_0(g^2)\right) + \mu_0\left(b_0^2\left[\frac{1}{2}(g^2)'' - gg''\right]\right) = \mu_0\left((b_0 g')^2\right),$$

where we have used that $\mu_0(\mathcal{L}_0(g^2)) = 0$, see e.g. Hansen and Scheinkman (1995), p. 774. □

Proof of Theorem 4.2. Under the conditions of theorem, the function κ is 1-1, and κ^{-1} is continuous. By Lemma 3.1, $V_n(f) \xrightarrow{\mathbb{P}_0} \kappa(\theta_0)$ as $n \to \infty$. We have assumed that $\theta_0 \in \operatorname{int}\Theta$, so $\kappa(\theta_0) \in \operatorname{int}\kappa(\Theta)$, and hence $\mathbb{P}_0(V_n(f) \in \kappa(\Theta)) \to 1$ as $n \to \infty$.

When $V_n(f) \in \kappa(\Theta)$, $\hat{\theta}_n = \kappa^{-1}(V_n(f))$ is the unique G_n-estimator. When $V_n(f) \notin \kappa(\Theta)$, we set $\hat{\theta}_n := \theta^*$ for some $\theta^* \in \Theta$. Then $\hat{\theta}_n \xrightarrow{\mathbb{P}_0} \theta_0$ as $n \to \infty$, and by a Taylor expansion

$$\sqrt{n\Delta_n}\left(\hat{\theta}_n - \theta_0\right) = \partial_\theta \kappa(\theta_0)\sqrt{n\Delta_n} V_n(f^*) + o_{\mathbb{P}_0}(1),$$

so (19) follows from Proposition 3.4. □

Proof of Lemma 4.4. To simplify the presentation, we define

$$H_n(\theta) = \frac{1}{n\Delta_n}\sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \tag{39}$$

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

$$g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) = f(X_{t_i^n}) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}) \tag{40}$$

$$g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) = f(X_{t_{i-1}^n}) g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta). \tag{41}$$

As a first step we verify the expansion (21) of $\breve{a}_n(\theta)$ in powers of Δ_n. By Lemma B.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

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

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

$$\breve{a}_n(\theta)_1 = \frac{\mathbb{E}_\theta\left[f(X_0)f(X_{\Delta_n})\right] - [\mu_\theta(f)]^2}{\mathbb{V}\mathrm{ar}_\theta f(X_0)} = 1 + \Delta_n K_f(\theta) + \Delta_n^2 R(\Delta_n; \theta), \tag{42}$$

and, as a consequence,

$$\breve{a}_n(\theta)_0 = -\Delta_n K_f(\theta)\mu_\theta(f) + \Delta_n^2 R(\Delta_n; \theta). \tag{43}$$

This expansion of $\breve{a}_n(\theta)$ together with

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

imply that

$$\begin{aligned} & \mathbb{E}_0\left[g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \;\middle|\; \mathcal{F}_{i-1}^n\right] \\ = \; & \mathbb{E}_0\left(f(X_{t_i^n}) \;\middle|\; \mathcal{F}_{i-1}^n\right) - \breve{a}_n(\theta)_0 - \breve{a}_n(\theta)_1 f(X_{t_{i-1}^n}) \\ = \; & \Delta_n\left(\mathcal{L}_0 f(X_{t_{i-1}^n}) + K_f(\theta)\left[\mu_\theta(f) - f(X_{t_{i-1}^n})\right]\right) + \Delta_n^2 R_0(\Delta_n, X_{t_{i-1}^n}; \theta). \end{aligned} \tag{44}$$

Hence, by Lemma 3.1,

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

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

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

$$\frac{1}{n^2\Delta_n^2}\sum_{i=1}^n \mathbb{E}_0\left[g_1^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\;\Big|\;\mathcal{F}_{i-1}^n\right] = o_{\mathbb{P}_0}(1). \tag{45}$$

From the expansions (42) and (43), 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

$$\begin{aligned}
&g_1^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) = \qquad (46)\\
&\quad\left[f(X_{t_i^n}) - f(X_{t_{i-1}^n})\right]^2 + 2\left[f(X_{t_i^n}) - f(X_{t_{i-1}^n})\right]\Delta_n R(\Delta_n, X_{t_{i-1}^n};\theta) + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n};\theta).
\end{aligned}$$

Lemma B.1 implies that

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

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

$$\frac{1}{n^2\Delta_n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n};\theta)\mathbb{E}_0\left[|f(X_{t_i^n}) - f(X_{t_{i-1}^n})|\;\Big|\;\mathcal{F}_{i-1}^n\right] = \frac{\Delta_n^{1/2}}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R_0(\Delta_n, X_{t_{i-1}^n};\theta) \xrightarrow{\mathbb{P}_0} 0,$$

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 (45). Thus, by Lemma 9 in Genon-Catalot and Jacod (1993),

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

Similarly for $g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)$, it follows easily from (44) that

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

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

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

so the first conclusion of the lemma follows from Lemma 9 in Genon-Catalot and Jacod (1993).

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 implies

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

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 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 expansion (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 under the regularity assumption (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 a compact subset $\mathcal{M} \subseteq \Theta$, note that

$$\|\partial_{\theta^T} H_n(\theta) - Z(f)A(\theta)\| \le \|Z_n(f)[A_n(\theta) - A(\theta)]\| + \|[Z_n(f) - Z(f)]A(\theta)\|$$

and, in particular,

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

By continuity of norms, $\|Z_n(f)\| \xrightarrow{\mathbb{P}_0} \|Z(f)\|$ and $\|Z_n(f) - Z(f)\| = o_{\mathbb{P}_0}(1)$, so (25) follows by observing that

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

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

Proof of Theorem 4.5. We continue with the notation (39)-(41) introduced above. Existence of a consistent sequence of G_n-estimators $(\hat{\theta}_n)$ follows from Theorem 2.5 in Jacod and Sørensen (2018), because the conclusions of Lemma 4.4 and the assumption that $W(\theta_0)$ is non-singular imply Condition 2.2 in Jacod and Sørensen (2018). The uniqueness result follows from Theorem 2.7 in Jacod and Sørensen (2018) under the identifiability condition $\gamma(\theta_0;\theta) \neq 0$ for $\theta \neq \theta_0$. The function $\theta \mapsto \gamma(\theta_0;\theta)$ is called $G(\theta)$ in Jacod and Sørensen (2018) and is necessarily continuous.

Asymptotic normality when $n\Delta_n^3 \to 0$ follows from Theorem 2.11 in Jacod and Sørensen (2018). We only need to check that

$$\sqrt{n\Delta_n} H_n(\theta_0) \xrightarrow{\mathscr{D}_0} \mathcal{N}_2(0, \mathcal{V}_0(f)). \tag{47}$$

We apply the Cramér-Wold device to prove this weak convergence result, i.e. we must prove that for all $c_1, c_2 \in \mathbb{R}$

$$\begin{aligned} C_n &= c_1 \cdot \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) + c_2 \cdot \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \\ &\xrightarrow{\mathscr{D}_0} \mathcal{N}\left(0, \mu_0\left(\left[\partial_x U_0(c_1 f_1^* + c_2 f_2^*) + c_2 f f'\right]^2 b^2(\,\cdot\,;\theta_0)\right)\right) \end{aligned} \tag{48}$$

Reusing the expansions (42) and (43), 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 4.3. Hence,

$$\begin{aligned} &\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \\ &= \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n \left[f(X_{t_i^n}) - f(X_{t_{i-1}^n})\right] + \sqrt{n\Delta_n} \cdot V_n(f_1^*) + (n\Delta_n^3)^{1/2} \cdot \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}; \theta_0) \\ &= \sqrt{n\Delta_n} \cdot V_n(f_1^*) + o_{\mathbb{P}_0}(1), \end{aligned}$$

because the first term in the expansion is a telescoping sum. Note that asymptotic normality for the first coordinate of the estimating function follows from Proposition 3.4. However, to obtain joint weak convegence, we need to consider the second coordinate too, which requires 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

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

and, hence, by applying the expansions (42) and (43) 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 (36) 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,

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

where $f^* = c_1 f_1^* + c_2 f_2^*$.

To gather the non-negligible terms, we argue as in (38) that

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

which, in turn, yields the stochastic integral representation

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

At this point, we can apply the central limit theorem for martingale difference arrays; see e.g. Hall and Heyde (1980) or Häusler and Luschgy (2015). To shorten notation in the following, we define

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

and

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

First, by the conditional Itô isometry, Tonelli's theorem and Lemma B.2,

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

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 B.2, respectively, imply that

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

so based on the inequality

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

we conclude that

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

Now the martingale central limit theorem for triangular arrays implies (48), so (47) follows by the Cramér-Wold device. The alternative expressions for the matrix $\mathcal{V}_0(f)$ follows because by Proposition 3.4 $\mu_0\left([\partial_x U_0(g)b(\,\cdot\,;\theta_0)]^2\right) = 2\mu_0\left(gU_0(g)\right)$ for $g \in \mathscr{H}_0^2$, and because with $g_i(x) = U_0(f_i^*)$ and $b_0(x) = b(x;\theta_0)$ it follows from Proposition 3.3 that

$$\mu_0(f_1^*g_2 + f_2^*g_1) = -\mu_0\left(\mathcal{L}_0(g_1g_2)\right) + \mu_0\left(b_0^2g_1'g_2'\right) = \mu_0\left(b_0^2g_1'g_2'\right),$$

where we have used that $\mu_0(\mathcal{L}_0(g_1g_2)) = 0$, see e.g. Hansen and Scheinkman (1995), p. 774.

□

Proof of Proposition 5.1. By the Cauchy-Schwarz inequality and the inequality (15)

$$|\mu_0\left(f^*U_0(f^*)\right)| \leq \|f^*\|_2\,\|U_0(f^*)\|_2 \leq \frac{\|f^*\|_2^2}{\lambda_0},$$

where $\lambda_0 > 0$ denotes the spectral gap of (X_t) under $\mathbb{P}_0$. Hence,

$$\mathrm{AVAR}(\hat{\theta}_n) = \frac{2\,\mu_0\left(f^*U_0(f^*)\right)}{[\partial_\theta\mu_{\theta_0}(f)]^2} \leq \frac{2\,\mathbb{V}\mathrm{ar}_0 f(X_0)}{\lambda_0\,[\partial_\theta\mu_{\theta_0}(f)]^2}.$$

□

Appendix B: Moment expansions

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

Lemma B.1. *Let $f \in \mathcal{C}_p^1(S)$. For any $k \geq 1$, there exists a constant $C_{k,\theta} > 0$ such that*

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

For completeness, we give a rough proof of the following theorem.

Lemma B.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, 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 B.1. □