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

Nina Munkholt Jakobsen
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
munkholt@math.ku.dk

Michael Sørensen*
Department of Mathematical Sciences
University of Copenhagen
Universitetsparken 5
DK-2100 Copenhagen Ø
Denmark
michael@math.ku.dk

July 24, 2015

Abstract

Parametric estimation for diffusion processes is considered for high frequency observations over a fixed time interval. The processes solve stochastic differential equations with an unknown parameter in the diffusion coefficient. We find easily verified conditions on approximate martingale estimating functions under which estimators are consistent, rate optimal, and efficient under high frequency (in-fill) asymptotics. The asymptotic distributions of the estimators are shown to be normal variance-mixtures, where the mixing distribution generally depends on the full sample path of the diffusion process over the observation time interval. Utilising the concept of stable convergence, we also obtain the more easily applicable result that for a suitable data dependent normalisation, the estimators converge in distribution to a standard normal distribution. The theory is illustrated by a small simulation study comparing an efficient and a non-efficient estimating function.

Key words: Approximate martingale estimating functions, discrete time sampling of diffusions, in-fill asymptotics, normal variance-mixtures, optimal rate, random Fisher information, stable convergence, stochastic differential equation.

Running title: Efficient Estimation for High Frequency SDE Data.

1 Introduction

Diffusions given by stochastic differential equations find application in a number of fields where they are used to describe phenomena which evolve continuously in time. Some examples include agronomy (Pedersen, 2000), biology (Favetto and Samson, 2010), finance (Merton, 1971; Vasicek, 1977; Cox et al., 1985; Larsen and Sørensen, 2007) and neuroscience (Ditlevsen and Lansky, 2006; Picchini et al., 2008; Bibbona et al., 2010).

While the models have continuous-time dynamics, data are only observable in discrete time, thus creating a demand for statistical methods to analyse such data. With the exception of some simple cases, the likelihood function is not explicitly known, and a large variety of alternate estimation procedures have been proposed in the literature, see e.g. Sørensen (2004) and Kessler et al. (2012). Parametric methods include the following. Maximum likelihood-type estimation using, primarily, Gaussian approximations to the likelihood function was considered by Prakasa Rao (1983), Florens-Zmirou (1989), Yoshida (1992), Genon-Catalot and Jacod (1993), Kessler (1997), Jacod (2006), Gloter and Sørensen (2009) and Uchida and Yoshida (2013). Analytical expansions of the transition densities were investigated by Aït-Sahalia (2002, 2008) and Li (2013), while approximations to the score function were studied by Bibby and Sørensen (1995), Kessler and Sørensen (1999), Jacobsen (2001, 2002), Uchida (2004), and Sørensen (2010). Simulation-based likelihood methods were developed by Pedersen (1995), Roberts and Stramer (2001), Durham and Gallant (2002), Beskos et al. (2006, 2009), Golightly and Wilkinson (2006, 2008), Bladt and Sørensen (2014), and Bladt et al. (2015).

A large part of the parametric estimators proposed in the literature can be treated within the framework of approximate martingale estimating functions, see the review in Sørensen (2012). In this paper, we derive easily verified conditions on such estimating functions that imply rate optimality and efficiency under a high frequency asymptotic scenario and thus contribute to providing clarity and a systematic approach to this area of statistics.

Specifically, the paper concerns parametric estimation for stochastic differential equations of the form

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

where $(W_t)_{t\geq 0}$ is a standard Wiener process. The drift and diffusion coefficients a and b are known, deterministic functions, and θ is the unknown parameter to be estimated. For ease of exposition, X_t and θ are both assumed to be one-dimensional. The extension of our results to a multivariate parameter is straightforward, and it is expected that multivariate diffusions can be treated in a similar way. For $n \in \mathbb{N}$, we consider observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$ in the time interval $[0,1]$, at discrete, equidistant time-points $t_i^n = i/n$, $i = 0, 1, \ldots, n$. We investigate the high frequency scenario where $n \to \infty$. The choice of the time-interval $[0,1]$ is not restrictive since results generalise to other compact intervals by suitable rescaling of the drift and diffusion coefficients. The drift coefficient does not depend on the parameter, because parameters that appear only in the drift cannot be estimated consistently in our asymptotic scenario.

It was shown by Dohnal (1987) and Gobet (2001) that under the asymptotic scenario considered here, the model (1.1) is locally asymptotic mixed normal with rate $\sqrt{n}$ and random asymptotic Fisher information

$$\mathcal{I}(\theta) = 2 \int_0^1 \left(\frac{\partial_\theta b(X_s; \theta)}{b(X_s; \theta)} \right)^2 ds. \tag{1.2}$$

Thus, a consistent estimator $\hat{\theta}_n$ is rate optimal if $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to a non-degenerate random variable as $n \to \infty$, where θ_0 is the true parameter value. The estimator is efficient if the limit may be written on the form $\mathcal{I}(\theta_0)^{-1/2} Z$, where Z is standard normal distributed and independent of $\mathcal{I}(\theta_0)$. The concept of local asymptotic mixed normality was introduced by Jeganathan (1982), and is discussed in e.g. Le Cam and Yang (2000, Chapter 6) and Jacod (2010).

Estimation for the model (1.1) under the high frequency asymptotic scenario described above was considered by Genon-Catalot and Jacod (1993, 1994). These authors proposed estimators based on a class of constrast functions that were only allowed to depend on the observations through $b^2(X_{t_{i-1}^n}; \theta)$ and $\Delta_n^{-1/2}(X_{t_i^n} - X_{t_{i-1}^n})$. The estimators were shown to be rate optimal, and a condition for efficiency was given.

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

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

with $\Delta_n = 1/n$, where the real-valued function $g(t, y, x; \theta)$ satisfies that $\mathbb{E}_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n})$ is of order Δ_n^κ for some $\kappa \geq 2$. Estimators are obtained as solutions to the estimating equation $G_n(\theta) = 0$ and are referred to as G_n-estimators.

This class of estimating functions was also studied by Sørensen (2010), who considered high frequency observations in an increasing time interval for a model like (1.1) where also the drift coefficient depends on a parameter. Specifically, the observation times were $t_i^n = i\Delta_n$ with $\Delta_n \to 0$ and $n\Delta_n \to \infty$. Simple conditions on g for rate optimality and efficiency were found under the infinite horizon high frequency asymptotics. To some extent, the methods of proof in the present paper are similar to those in Sørensen (2010). However, while ergodicity of the diffusion process played a central role in Sørensen (2010), this property is not needed here. Another important difference is that here expansions of a higher order are needed, which complicates the proofs considerably. Furthermore, here a more complicated version of the central limit theorem for martingales is required, and we need the concept of *stable* convergence in distribution, in order to obtain practically applicable convergence results.

First, we establish results on existence and uniqueness of consistent G_n-estimators. We show that $\sqrt{n}(\hat{\theta}_n - \theta_0)$ converges in distribution to a normal variance-mixture, which implies rate optimality. The limit distribution may be represented by the product $W(\theta_0)Z$ of independent random variables, where Z is standard normal distributed. The random variable $W(\theta_0)$ is generally non-degenerate, and depends on the entire path of the diffusion process over the time-interval $[0, 1]$. Normal variance-mixtures were also obtained as the asymptotic distributions of the estimators of Genon-Catalot and Jacod (1993). These distributions appear as limit distributions in comparable non-parametric settings as well, e.g. when estimating integrated volatility (Jacod and Protter, 1998; Mykland and Zhang, 2006) or the squared diffusion coefficient (Florens-Zmirou, 1993; Jacod, 2000).

Rate optimality is ensured by the condition that

$$\partial_y g(0, x, x; \theta) = 0 \tag{1.3}$$

for all x in the state space of the diffusion process, and all parameter values θ. Here $\partial_y g(0, x, x; \theta)$ denotes the first derivative of $g(0, y, x; \theta)$ with respect to y evaluated in $y = x$. The same condition was found in Sørensen (2010) for rate optimality of an estimator of the parameter in the diffusion coefficient and as one of the conditions for small Δ-optimality; see Jacobsen (2001, 2002).

Due to its dependence on $(X_s)_{s\in[0,1]}$, the limit distribution is difficult to use for statistical applications, such as constructing confidence intervals and test statistics. Therefore, we construct a statistic $\widehat{W}_n$ that converges in probability to $W(\theta_0)$. Using the stable convergence in distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ towards $W(\theta_0)Z$, we derive the more easily applicable result that $\sqrt{n}\,\widehat{W}_n^{-1}(\hat{\theta}_n - \theta_0)$ converges in distribution to a standard normal distribution.

The additional condition that

$$\partial_y^2 g(0, x, x; \theta) = K_\theta \frac{\partial_\theta b^2(x; \theta)}{b^4(x; \theta)} \tag{1.4}$$

($K_\theta \neq 0$) for all x in the state space, and all parameter values θ, ensures efficiency of G_n-estimators. The same condition was obtained by Sørensen (2010) in his infinite horizon scenario for efficiency of estimators of parameters in the diffusion coefficient. It is also identical to a condition given by Jacobsen (2002) for small Δ-optimality. The identity of the conditions implies that examples of approximate martingale estimating functions that are rate optimal and efficient in our asymptotic scenario may be found in Jacobsen (2002) and Sørensen (2010). In particular, estimating functions that are optimal in the sense of Godambe and Heyde (1987) are rate optimal and efficient under weak regularity conditions.

The paper is structured as follows: Section 2 presents definitions, notation and terminology used throughout the paper, as well as the main assumptions. Section 3 states and discusses our main results, while Section 4 presents a simulation study illustrating the results. Section 5 contains main lemmas used to prove the main theorem and proofs of the main theorem and of the lemmas. Appendix A consists of auxiliary technical results, some of them with proofs.

2 Preliminaries

2.1 Model and Observations

Let $(\Omega, \mathcal{F})$ be a measurable space supporting a real-valued random variable U, and an independent standard Wiener process $\mathbf{W} = (W_t)_{t\geq 0}$. Let $(\mathcal{F}_t)_{t\geq 0}$ denote the filtration generated by U and $\mathbf{W}$.

Consider the stochastic differential equation

$$dX_t = a(X_t)\,dt + b(X_t;\theta)\,dW_t\,, \quad X_0 = U\,, \tag{2.1}$$

for $\theta \in \Theta \subseteq \mathbb{R}$. The state space of the solution is assumed to be an open interval $\mathcal{X} \subseteq \mathbb{R}$, and the drift and diffusion coefficients, $a : \mathcal{X} \to \mathbb{R}$ and $b : \mathcal{X} \times \Theta \to \mathbb{R}$, are assumed to be known, deterministic functions. Let $(\mathbb{P}_\theta)_{\theta\in\Theta}$ be a family of probability measures on $(\Omega, \mathcal{F})$ such that $\mathbf{X} = (X_t)_{t\geq 0}$ solves (2.1) under $\mathbb{P}_\theta$, and let $\mathbb{E}_\theta$ denote expectation under $\mathbb{P}_\theta$.

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

2.2 Polynomial Growth

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

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

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

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

Definition 2.1. $C^{\text{pol}}_{p,q,r}([0,1]\times\mathcal{X}^2\times\Theta)$ denotes the class of continuous, real-valued functions $f(t,y,x;\theta)$ which satisfy that

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

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

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

Note that in Definition 2.1, differentiability of f with respect to x is never required.

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

2.3 Approximate Martingale Estimating Functions

Definition 2.2. Let $g(t, y, x; \theta)$ be a real-valued function defined on $[0, 1]\times\mathcal{X}^2\times\Theta$. Suppose the existence of a constant $\kappa \geq 2$, such that for all $n \in \mathbb{N}$, $i = 1, \ldots, n$, $\theta \in \Theta$,

$$\mathbb{E}_\theta \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n} \right) = \Delta_n^\kappa R_\theta(\Delta_n, X_{t_{i-1}^n}) . \tag{2.2}$$

Then, the function

$$G_n(\theta) = \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \tag{2.3}$$

is called an *approximate martingale estimating function*. In particular, when (2.2) is satisfied with $R_\theta(t, x) \equiv 0$, (2.3) is referred to as a *martingale estimating function*. $\diamond$

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

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

is a zero-mean, real-valued $(\mathcal{G}_{n,i})_{1\leq i\leq n}$-martingale under $\mathbb{P}_\theta$ for each $n \in \mathbb{N}$. The score function of the observations $(X_{t_0^n}, X_{t_1^n}, \ldots, X_{t_n^n})$ is under weak regularity conditions a martingale estimating function, and an approximate martingale estimating function can be viewed as an approximation to the score function.

A G_n*-estimator* $\hat{\theta}_n$ is essentially obtained as a solution to the estimating equation $G_n(\theta) = 0$. A more precise definition is given in the following Definition 2.3. Here we make the ω-dependence explicit by writing $G_n(\theta, \omega)$ and $\hat{\theta}_n(\omega)$.

Definition 2.3. Let $G_n(\theta, \omega)$ be an approximate martingale estimating function as defined in Definition 2.2. Put $\Theta_\infty = \Theta \cup \{\infty\}$ and let

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

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

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

2.4 Assumptions

We make the following assumptions about the stochastic differential equation.

Assumption 2.4. *The parameter set Θ is a non-empty, open subset of $\mathbb{R}$. Under the probability measure $\mathbb{P}_\theta$, the continuous, $(\mathcal{F}_t)_{t\geq 0}$-adapted Markov process $\mathbf{X} = (X_t)_{t\geq 0}$ solves a stochastic differential equation of the form (2.1), the coefficients of which satisfy that*

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

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

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

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

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

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

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

$\diamond$

The global Lipschitz condition, Assumption 2.4.(ii), ensures that a unique solution exists such that **X** is well-defined. Assumption 2.4 is very similar to the corresponding Condition 2.1 of Sørensen (2010). However, an important difference is that in the current paper, **X** is not required to be ergodic. Here, law of large numbers-type results are replaced by what is, in essence, the convergence of Riemann sums.

We make the following assumptions about the estimating function.

Assumption 2.5. *The function $g(t, y, x; \theta)$ satisfies (2.2) for some $\kappa \geq 2$, thus defining an approximate martingale estimating function by (2.3). Moreover,*

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

and the following holds for all $\theta \in \Theta$.

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

(ii) The expansion

$$\begin{aligned} g(\Delta, y, x; \theta) = g(0, y, x; \theta) &+ \Delta g^{(1)}(y, x; \theta) + \tfrac{1}{2}\Delta^2 g^{(2)}(y, x; \theta) + \tfrac{1}{6}\Delta^3 g^{(3)}(y, x; \theta) \\ &+ \Delta^4 R(\Delta, y, x; \theta) \end{aligned} \tag{2.4}$$

holds for all $\Delta \in [0, 1]$ and $x, y \in \mathcal{X}$, where $g^{(j)}(y, x; \theta)$ denotes the j'th partial derivative of $g(t, y, x; \theta)$ with respect to t, evaluated in $t = 0$.

$\diamond$

Assumption 2.5.(i) was referred to by Sørensen (2010) as *Jacobsen's condition*, as it is one of the conditions for small Δ-optimality in the sense of Jacobsen (2001), see Jacobsen (2002). The assumption ensures rate optimality of the estimators in this paper, and of the estimators of the parameters in the diffusion coefficient in Sørensen (2010).

The assumptions of polynomial growth and existence and boundedness of all moments serve to simplify the exposition and proofs, and could be relaxed.

2.5 The Infinitesimal Generator

For $\lambda \in \Theta$, the infinitesimal generator $\mathcal{L}_\lambda$ is defined for all functions $f(y) \in C_2^{\text{pol}}(\mathcal{X})$ by

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

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

$$\mathcal{L}_\lambda f(t,y,x;\theta) = a(y)\partial_y f(t,y,x;\theta) + \tfrac{1}{2}b^2(y;\lambda)\partial_y^2 f(t,y,x;\theta)\,. \tag{2.5}$$

Often, the notation $\mathcal{L}_\lambda f(t,y,x;\theta) = \mathcal{L}_\lambda(f(t;\theta))(y,x)$ is used, so e.g. $\mathcal{L}_\lambda(f(0;\theta))(x,x)$ means $\mathcal{L}_\lambda f(0,y,x;\theta)$ evaluated in $y = x$. In this paper the infinitesimal generator is particularly useful because of the following result.

Lemma 2.6. *Suppose that Assumption 2.4 holds, and that for some $k \in \mathbb{N}_0$,*

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

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

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

where, furthermore,

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

◇

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

For concrete models, Lemma 2.6 is useful for verifying the approximate martingale property (2.2) and for creating approximate martingale estimating functions. In combination with (2.2), the lemma is key to proving the following Lemma 2.7, which reveals two important properties of approximate martingale estimating functions.

Lemma 2.7. *Suppose that Assumptions 2.4 and 2.5 hold. Then*

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

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

Lemma 2.7 corresponds to Lemma 2.3 of Sørensen (2010), to which we refer for details on the proof.

3 Main Results

Section 3.1 presents the main theorem of this paper, which establishes existence, uniqueness and asymptotic distribution results for rate optimal estimators based on approximate martingale estimating functions. In Section 3.2 a condition is given, which ensures that the rate optimal estimators are also efficient, and efficient estimators are discussed.

3.1 Main Theorem

The final assumption needed for the main theorem is as follows.

Assumption 3.1. *The following holds $\mathbb{P}_\theta$-almost surely for all $\theta \in \Theta$.*

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

$$\int_0^1 (b^2(X_s; \theta) - b^2(X_s; \lambda))\partial_y^2 g(0, X_s, X_s; \lambda)\, ds \neq 0\,,$$

(ii)

$$\int_0^1 \partial_\theta b^2(X_s; \theta)\partial_y^2 g(0, X_s, X_s; \theta)\, ds \neq 0\,,$$

(iii)

$$\int_0^1 b^4(X_s; \theta)\left(\partial_y^2 g(0, X_s, X_s; \theta)\right)^2\, ds \neq 0\,.$$

◇

Assumption 3.1 can be difficult to check in practice because it involves the full sample path of $\mathbf{X}$ over the interval $[0, 1]$. It requires, in particular, that for all $\theta \in \Theta$, with $\mathbb{P}_\theta$-probability one, $t \mapsto b^2(X_t; \theta) - b^2(X_t; \lambda)$ is not Lebesgue-almost surely zero when $\lambda \neq \theta$. As noted by Genon-Catalot and Jacod (1993), this requirement holds true (by the continuity of the function) if, for example, $X_0 = U$ is degenerate at x_0, and $b^2(x_0; \theta) \neq b^2(x_0; \lambda)$ for all $\theta \neq \lambda$.

For an efficient estimating function, Assumption 3.1 reduces to conditions on $\mathbf{X}$ with no further conditions on the estimating function, see the next section. Specifically, the conditions involve only the squared diffusion coefficient $b^2(x; \theta)$ and its derivative $\partial_\theta b^2$.

Theorem 3.2. *Suppose that Assumptions 2.4, 2.5 and 3.1 hold. Then,*

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

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

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} W(\theta_0)Z\,. \tag{3.1}$$

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

$$W(\theta_0) = \frac{\left(\int_0^1 \frac{1}{2} b^4(X_s;\theta_0)\left(\partial_y^2 g(0,X_s,X_s;\theta_0)\right)^2 \, ds\right)^{1/2}}{\int_0^1 \frac{1}{2}\partial_\theta b^2(X_s;\theta_0)\partial_y^2 g(0,X_s,X_s;\theta_0)\, ds}\,. \tag{3.2}$$

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

$$\widehat{W}_n = -\frac{\left(\frac{1}{\Delta_n}\sum_{i=1}^n g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\theta}_n)\right)^{1/2}}{\sum_{i=1}^n \partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\hat{\theta}_n)} \tag{3.3}$$

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

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

where Z is standard normally distributed.

◇

The proof of Theorem 3.2 is given in Section 5.1.

Dohnal (1987) and Gobet (2001) showed local asymptotic mixed normality with rate $\sqrt{n}$, so Theorem 3.2 establishes rate optimality of G_n-estimators.

Observe that the limit distribution in Theorem 3.2.(ii) generally depends on not only the unknown parameter θ_0, but also on the concrete realisation of the sample path $t \mapsto X_t$ over $[0,1]$, which is only partially observed. In contrast, Theorem 3.2.(iii) yields a limit distribution which is of more use in practical applications.

3.2 Efficiency

Under the assumptions of Theorem 3.2, the following additional condition ensures efficiency of a consistent G_n-estimator.

Assumption 3.3. *Suppose that for each $\theta \in \Theta$, there exists a constant $K_\theta \neq 0$ such that for all $x \in \mathcal{X}$,*

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

$\diamond$

Dohnal (1987) and Gobet (2001) showed that the local asymptotic mixed normality property holds within the framework considered here with random Fisher information $\mathcal{I}(\theta_0)$ given by (1.2). Thus, a G_n-estimator $\hat{\theta}_n$ is efficient if (3.1) holds with $W(\theta_0) = \mathcal{I}(\theta_0)^{-1/2}$, so the following Corollary 3.4 may easily be verified.

Corollary 3.4. *Suppose that the assumptions of Theorem 3.2 and Assumption 3.3 hold. Then, any consistent G_n-estimator is also efficient.* $\diamond$

It follows from Theorem 3.2 and Lemma 5.1 that if Assumption 3.3 holds, and if G_n is normalized such that $K_\theta = 1$, then

$$\sqrt{n}\, \widehat{\mathcal{I}}_n^{\frac{1}{2}} (\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0,1) ,$$

where

$$\widehat{\mathcal{I}}_n = \frac{1}{\Delta_n} \sum_{i=1}^n g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n).$$

It was noted in Section 2.3 that not necessarily all versions of a particular estimating function satisfy the conditions of this paper, even though we obtain the same estimator. Thus, an estimating function is said to be efficient, if there exists a version which satisfies the conditions of Corollary 3.4. The same goes for rate optimality.

Assumption 3.3 is identical to the condition for efficiency of estimators of parameters in the diffusion coefficient in Sørensen (2010), and to one of the conditions for small Δ-optimality in Jacobsen (2002).

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

$$\bar{g}(t, y, x; \theta) = \frac{\partial_\theta b^2(x;\theta)}{b^4(x;\theta)} \left((y-x)^2 - t b^2(x;\theta) \right) \tag{3.4}$$

yields an example of an efficient estimating function. The approximate martingale property (2.2) can be verified by Lemma 2.6.

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

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

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

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

and $\partial_\theta \bar{u}(t, y, x; \theta) = -\bar{g}(t, y, x; \theta)$. In other words, it corresponds to a version of the efficient approximate martingale estimating function given by (3.4). The same contrast function was considered by Uchida and Yoshida (2013) in the framework of a more general class of stochastic differential equations.

A problem of considerable practical interest is how to construct estimating functions that are rate optimal and efficient, i.e. estimating functions satisfying Assumptions 2.5.(i) and 3.3. Being the same as the conditions for small Δ-optimality, the assumptions are, for example, satisfied for martingale estimating functions constructed by Jacobsen (2002).

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

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

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

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

satisfies Assumptions 2.5.(i) and 3.3, and is thus rate optimal and efficient. As there is a simple formula for $\bar{a}$ (see Section 1.11.1 of Sørensen (2012)), this provides a way of constructing a large number of efficient estimating functions. The result also holds if $\phi_\theta^t f(x; \theta)$ and the conditional moments in the formula for $\bar{a}$ are approximated suitably by the help of Lemma 2.6.

Remark 3.5. Suppose for a moment that the diffusion coefficient of (2.1) has the form $b^2(x; \theta) = h(x)k(\theta)$ for strictly positive functions h and k, with Assumption 2.4 satisfied. This holds true, e.g., for a number of Pearson diffusions, including the (stationary) Ornstein-Uhlenbeck and square root processes. (See Forman and Sørensen (2008) for more on Pearson diffusions.) Then $\mathcal{I}(\theta_0) = \partial_\theta k(\theta_0)^2 / (2k^2(\theta_0))$. In this case, under the assumptions of Corollary 3.4, an efficient G_n-estimator $\hat{\theta}_n$ satisfies that $\sqrt{n}(\hat{\theta}_n - \theta_0) \longrightarrow Y$ in distribution where Y is normally distributed with mean zero and variance $2k^2(\theta_0)/\partial_\theta k(\theta_0)^2$, i.e. the limit distribution is not a normal variance-mixture depending on $(X_t)_{t \in [0,1]}$. Note also that when $b^2(x; \theta) = h(x)k(\theta)$ and Assumption 3.3 holds, then Assumption 3.1 is satisfied when e.g. $\partial_\theta k(\theta) > 0$ or $\partial_\theta k(\theta) < 0$. ◦

4 Simulation study

This section presents a simulation study illustrating the theory in the previous section. An efficient and an inefficient estimating function are compared for a model for which the limit distributions of the consistent estimators are non-degenerate normal variance-mixtures.

Consider the stochastic differential equation

$$dX_t = -2X_t\,dt + (\theta + X_t^2)^{-1/2}\,dW_t \tag{4.1}$$

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

$$\mu_\theta(x) = \exp\left(-2\theta x^2 - x^4\right)\left(\theta + x^2\right), \quad x \in \mathbb{R}, \tag{4.2}$$

with respect to Lebesgue measure. When $\mathbf{X}$ is stationary, the process satisfies Assumption 2.4. We compare the two estimating functions given by

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

where

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

Both g and h satisfy Assumptions 2.5 and 3.1, and g is the efficient function (3.4), while h is not efficient.

Let $W_G(\theta_0)$ and $W_H(\theta_0)$ be given by (3.2), that is

$$W_G(\theta_0) = \left(\frac{1}{2}\int_0^1 \frac{1}{(\theta_0 + X_s^2)^2}\,ds\right)^{-1/2} \quad \text{and} \quad W_H(\theta_0) = \frac{\left(\int_0^1 2(\theta_0 + X_s^2)^{18}\,ds\right)^{1/2}}{\int_0^1 (\theta_0 + X_s^2)^8\,ds}\,. \tag{4.3}$$

Numerical calculations and simulations were done in R 3.1.2 (R Core Team, 2014). First, $m = 10^4$ trajectories of the process $\mathbf{X}$ given by (4.1) were simulated over the time-interval $[0, 1]$ with $\theta_0 = 1$, each with sample size $n = 10^4$. These simulations were performed using the R-package *sde* (Iacus, 2014). For each trajectory, the initial value X_0 was obtained from the invariant distribution of $\mathbf{X}$ by inverse transform sampling, using a quantile function based on (4.2), and calculated by numerical procedures in R. For $n = 10^3$ and $n = 10^4$, let $\hat{\theta}_{G,n}$ and $\hat{\theta}_{H,n}$ denote estimates of θ obtained by solving the equations $G_n(\theta) = 0$ and $H_n(\theta) = 0$ numerically, on the interval $[0.01, 1, 99]$. Using these estimates, $\widehat{W}_{G,n}$ and $\widehat{W}_{H,n}$ are calculated by (3.3). For $n = 10^3$, $\hat{\theta}_{H,n}$ and thus also $\widehat{W}_{H,n}$, could not be computed for 44 of the $m = 10^4$ sample paths. For $n = 10^4$, and for the efficient estimator $\hat{\theta}_{G,n}$ there were no problems.

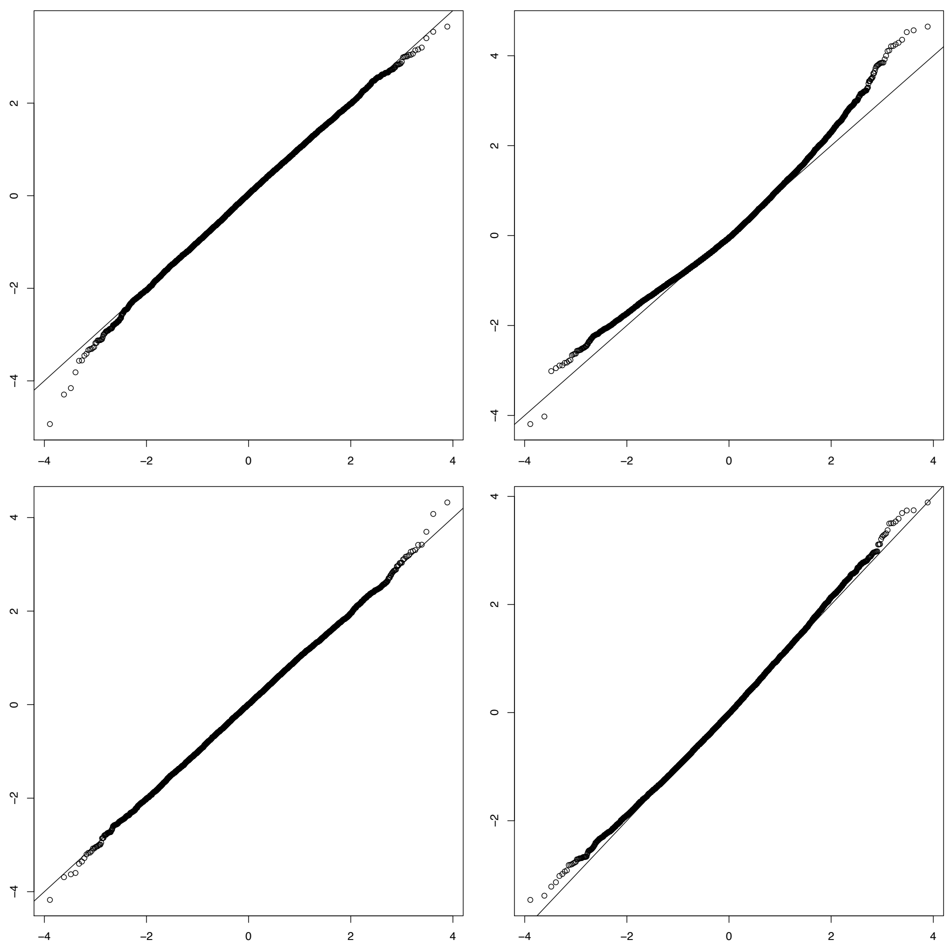

Figure 1: QQ-plots comparing $\widehat{Z}_{G,n}$ (left) and $\widehat{Z}_{H,n}$ (right) to the $\mathcal{N}(0,1)$ distribution for $n = 10^3$ (above) and $n = 10^4$ (below).

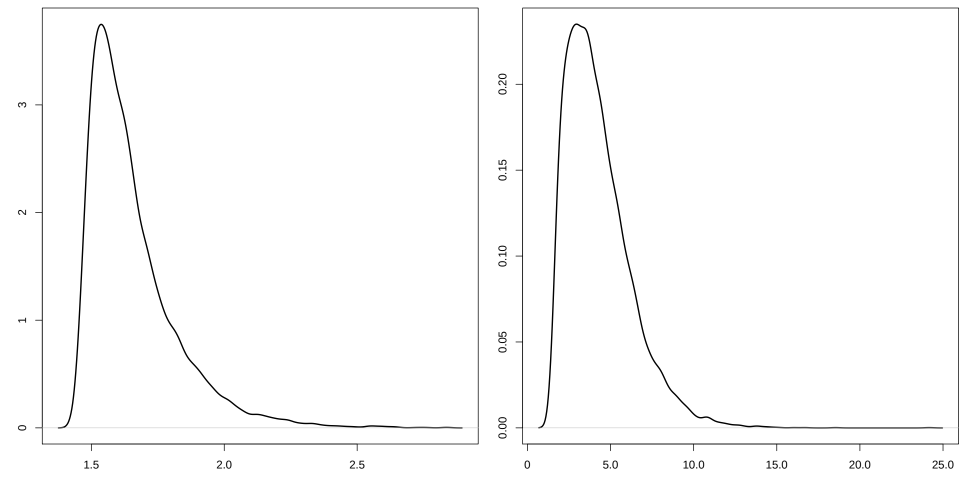

Figure 2: Approximation to the densities of $W_G(\theta_0)$ (left) and $W_H(\theta_0)$ (right) based on $\widetilde{W}_G$ and $\widetilde{W}_H$.

Figure 1 shows QQ-plots of

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

compared with a standard normal distribution, for $n = 10^3$ and $n = 10^4$ respectively. These QQ-plots suggest that, as n goes to infinity, the asymptotic distribution in Theorem 3.2.(iii) becomes a good approximation faster in the efficient case than in the inefficient case.

Inserting $\theta_0 = 1$ into (4.3), the intergrals in these expressions may be approximated by Riemann sums, using each of the simulated trajectories of $\mathbf{X}$ (with $n = 10^4$ for maximal accuracy). This method yields a second set of approximations $\widetilde{W}_G$ and $\widetilde{W}_H$ to the realisations of the random variables $W_G(\theta_0)$ and $W_H(\theta_0)$, presumed to be more accurate than $\widehat{W}_{G,10^4}$ and $\widehat{W}_{H,10^4}$ as they utilise the true parameter value. The *density* function in R was used (with default arguments) to compute an approximation to the densities of $W_G(\theta_0)$ and $W_H(\theta_0)$, using the approximate realisations $\widetilde{W}_G$ and $\widetilde{W}_H$.

It is seen from Figure 2 that the distribution of $W_H(\theta_0)$ is much more spread out than the distribution of $W_G(\theta_0)$. This corresponds well to the limit distribution in Theorem 3.2.(ii) being more spread out in the inefficient case than in the efficient case. Along the same lines, Figure 3 shows similarly computed densities based on $\sqrt{n}(\hat{\theta}_{G,n} - \theta_0)$ and $\sqrt{n}(\hat{\theta}_{H,n} - \theta_0)$ for $n = 10^4$, which may be considered approximations to the densities of the normal variance-mixture limit distributions in Theorem 3.2.(ii). These plots also illustrate that the limit distribution of the inefficient estimator is more spread out than that of the efficient estimator.

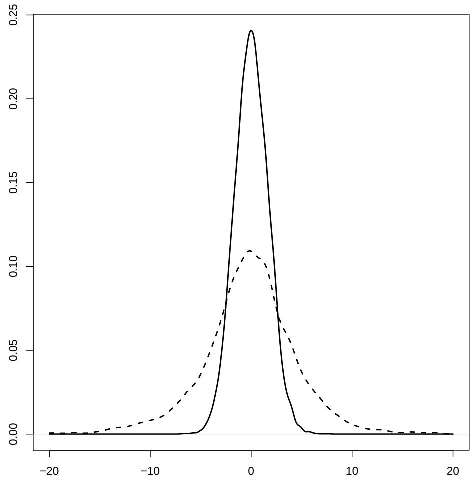

Figure 3: Estimated densities of $\sqrt{n}(\hat{\theta}_{G,n} - \theta_0)$ (solid curve) and $\sqrt{n}(\hat{\theta}_{H,n} - \theta_0)$ (dashed curve) for $n = 10^4$.

5 Proofs

Section 5.1 states three main lemmas needed to prove Theorem 3.2 which are followed by the proof of the theorem. Section 5.2 contains the proofs of the three lemmas.

5.1 Proof of the Main Theorem

In order to prove Theorem 3.2, we use the following lemmas, together with results from Jacod and Sørensen (2012), and Sørensen (2012, Section 1.10).

Lemma 5.1. *Suppose that Assumptions 2.4 and 2.5 hold. For $\theta \in \Theta$, let*

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

and

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

Then,

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

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

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \left| \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \right| \xrightarrow{\mathcal{P}} 0 \tag{5.1}$$

$$\frac{1}{\Delta_n} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right)^2 \xrightarrow{\mathcal{P}} 0 \tag{5.2}$$

$$\frac{1}{\Delta_n^2} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g^4(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} 0 \tag{5.3}$$

and

$$\frac{1}{\Delta_n} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g^2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} \tfrac{1}{2} \int_0^t b^4(X_s;\theta_0) \left(\partial_y^2 g(0, X_s, X_s; \theta_0) \right)^2 ds\,. \tag{5.4}$$

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

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

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

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

$\diamond$

Lemma 5.2. *Suppose that Assumptions 2.4 and 2.5 hold. Then, for all* $t > 0$,

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \mathbb{E}_{\theta_0} \left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} 0\,. \tag{5.5}$$

$\diamond$

Lemma 5.3. *Suppose that Assumptions 2.4 and 2.5 hold, and let*

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

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

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

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

We denote stable convergence in distribution under $\mathbb{P}_{\theta_0}$ as $n \to \infty$ by $\xrightarrow{\mathcal{D}_{st}}$.

Proof of Theorem 3.2. Let a compact, convex subset $K \subseteq \Theta$ with $\theta_0 \in \operatorname{int} K$ be given. The functions $G_n(\theta)$, $A(\theta, \theta_0)$, $B(\theta, \theta_0)$, and $C(\theta, \theta_0)$ were defined in Lemma 5.1.

By Lemma 5.1.(i) and (iii),

$$G_n(\theta_0) \xrightarrow{\mathcal{P}} 0 \quad \text{and} \quad \sup_{\theta \in K} |\partial_\theta G_n(\theta) - B(\theta, \theta_0)| \xrightarrow{\mathcal{P}} 0 \tag{5.6}$$

with $B(\theta_0; \theta_0) \neq 0$ by Assumption 3.1.(ii), so $G_n(\theta)$ satisfies the conditions of Theorem 1.58 in Sørensen (2012).

Now, we show (1.161) of Theorem 1.59 in Sørensen (2012). Let $\varepsilon > 0$ be given, and let $\bar{B}_\varepsilon(\theta_0)$ and $B_\varepsilon(\theta_0)$, respectively, denote closed and open balls in $\mathbb{R}$ with radius $\varepsilon > 0$, centered at θ_0. The compact set $K \backslash B_\varepsilon(\theta_0)$ does not contain θ_0, and so, by Assumption 3.1.(i), $A(\theta, \theta_0) \neq 0$ for all $\theta \in K \backslash B_\varepsilon(\theta_0)$ with probability one under $\mathbb{P}_{\theta_0}$.

Because

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

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

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

Consequently, by Theorem 1.59 in Sørensen (2012), for any G_n-estimator $\tilde{\theta}_n$,

$$\mathbb{P}_{\theta_0} \left(\tilde{\theta}_n \in K \backslash \bar{B}_\varepsilon(\theta_0) \right) \to 0 \quad \text{as} \quad n \to \infty \,. \tag{5.7}$$

for any $\varepsilon > 0$.

By Theorem 1.58 in Sørensen (2012), there exists a consistent G_n-estimator $\hat{\theta}_n$, which is eventually unique, in the sense that if $\bar{\theta}_n$ is another consistent G_n-estimator, then

$$\mathbb{P}_{\theta_0} \left(\hat{\theta}_n \neq \bar{\theta}_n \right) \to 0 \quad \text{as} \quad n \to \infty \,. \tag{5.8}$$

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

$$\mathbb{P}_{\theta_0} \left(\tilde{\theta}_n \in K \right) \to 1 \quad \text{as} \quad n \to \infty \,. \tag{5.9}$$

Combining (5.7) and (5.9), it follows that

$$\mathbb{P}_{\theta_0} \left(\tilde{\theta}_n \in \bar{B}_\varepsilon(\theta_0) \right) \to 1 \quad \text{as} \quad n \to \infty \,, \tag{5.10}$$

so $\tilde{\theta}_n$ is consistent. Using (5.8), Theorem 3.2.(i) follows.

To prove Theorem 3.2.(ii), recall that $\Delta_n = 1/n$, and observe that by Lemma 5.3,

$$\sqrt{n} G_n(\theta_0) \xrightarrow{\mathcal{D}_{st}} S(\theta_0) \tag{5.11}$$

where

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

and $\mathbf{B} = (B_s)_{s\in[0,1]}$ is a standard Wiener process, independent of U and $\mathbf{W}$. As $\mathbf{X}$ is then also independent of $\mathbf{B}$, $S(\theta_0)$ is equal in distribution to $C(\theta_0;\theta_0)^{1/2}Z$, where Z is standard normally distributed and independent of $(X_t)_{t\geq 0}$. Note that by Assumption 3.1.(iii), the distribution of $C(\theta_0;\theta_0)^{1/2}Z$ is non-degenerate.

Let $\hat{\theta}_n$ be a consistent G_n-estimator. By (5.6), (5.11) and properties of stable convergence (e.g. (2.3) in Jacod (1997)),

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

Stable convergence in distribution implies weak convergence, so an application of Theorem 1.60 in Sørensen (2012) yields

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} -B(\theta_0,\theta_0)^{-1}S(\theta_0)\,. \tag{5.12}$$

The limit is equal in distribution to $W(\theta_0)Z$, where $W(\theta_0) = -B(\theta_0,\theta_0)^{-1}C(\theta_0;\theta_0)^{1/2}$ and Z is standard normally distributed and independent of $W(\theta_0)$. This completes the proof of Theorem 3.2.(ii).

Finally, Lemma 2.14 in Jacod and Sørensen (2012) is used to write

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

where the last term goes to zero in probability under $\mathbb{P}_{\theta_0}$. By the stable continuous mapping theorem, (5.12) holds with stable convergence in distribution as well. Lemma 5.1.(iii) may be used to conclude that $\widehat{W}_n \xrightarrow{\mathcal{P}} W(\theta_0)$, so Theorem 3.2.(iii) follows from the stable version of (5.12) by application of standard results for stable convergence. □

5.2 Proofs of Main Lemmas

This section contains the proofs of Lemmas 5.1, 5.2 and 5.3 in Section 5.1. A number of technical results are utilised in the proofs, these results are summarised in Appendix A, some of them with a proof.

Proof of Lemma 5.1. First, note that for any $f(x;\theta) \in C^{\mathrm{pol}}_{0,0}(\mathcal{X}\times\Theta)$ and any compact, convex subset $K \subseteq \Theta$, there exist constants $C_K > 0$ such that

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

for all $s \in [0,1]$ and $\theta \in \operatorname{int} K$. With probability one under $\mathbb{P}_{\theta_0}$, for fixed ω, $C_K(1+|X_s(\omega)|^{C_K})$ is a continuous function and therefore Lebesgue-integrable over $[0,1]$. Using this method of constructing integrable upper bounds, Lemma 5.1.(i) follows by the usual results for continuity and differentiability of functions given by integrals.

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

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

proving (5.1) and (5.2). Furthermore, using (A.1) and (A.3),

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

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

$$G_n(\theta) - A(\theta; \theta_0) \xrightarrow{\mathcal{P}} 0\,. \tag{5.13}$$

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

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

and

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

completing the proof of Lemma 5.1.(ii) when $\theta = \theta_0$ is inserted, and yielding

$$G_n^{sq}(\theta) - C(\theta; \theta_0) \xrightarrow{\mathcal{P}} 0 \tag{5.14}$$

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

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

Thus, by Lemma A.1, also

$$\partial_\theta G_n(\theta) - B(\theta; \theta_0) \xrightarrow{\mathcal{P}} 0\,, \tag{5.15}$$

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

$$\begin{aligned}\mathbb{E}_{\theta_0}|(G_n(\theta) - A(\theta;\theta_0)) - (G_n(\theta') - A(\theta';\theta_0))|^{2m} &\le C_{m,K}\ |\theta - \theta'|^{2m}\\ \mathbb{E}_{\theta_0}|(\partial_\theta G_n(\theta) - B(\theta;\theta_0)) - (\partial_\theta G_n(\theta') - B(\theta';\theta_0))|^{2m} &\le C_{m,K}\ |\theta - \theta'|^{2m}\\ \mathbb{E}_{\theta_0}|(G_n^{sq}(\theta) - C(\theta;\theta_0)) - (G_n^{sq}(\theta') - C(\theta';\theta_0))|^{2m} &\le C_{m,K}\ |\theta - \theta'|^{2m}.\end{aligned} \tag{5.16}$$

By Lemma 5.1.(i), the functions $\theta \mapsto G_n(\theta) - A(\theta;\theta_0)$, $\theta \mapsto \partial_\theta G_n(\theta) - B(\theta;\theta_0)$ and $\theta \mapsto G_n^{sq}(\theta) - C(\theta, \theta_0)$ are continuous on Θ. Thus, using Lemma A.9 together with (5.13), (5.14), (5.15) and (5.16) completes the proof of Lemma 5.1.(iii). □

Proof of Lemma 5.2. The overall strategy in this proof is to expand the expression on the left-hand side of (5.5) in such a manner that all terms either converge to 0 by Lemma A.3, or are equal to 0 by the martingale properties of stochastic integral terms obtained by use of Itô's formula.

By Assumption 2.5 and Lemma 2.7, the formulae

$$\begin{aligned}g(0, y, x; \theta) &= \tfrac{1}{2}(y - x)^2 \partial_y^2 g(0, x, x; \theta) + (y - x)^3 R(y, x; \theta)\\ g^{(1)}(y, x; \theta) &= g^{(1)}(x, x; \theta) + (y - x) R(y, x; \theta)\end{aligned} \tag{5.17}$$

may be obtained. Using (2.4) and (5.17),

$$\begin{aligned}&\mathbb{E}_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\\ &\quad= \mathbb{E}_{\theta_0}\left(\tfrac{1}{2}(X_{t_i^n} - X_{t_{i-1}^n})^2 \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\\ &\qquad+ \mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n})^3 R(X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\\ &\qquad+ \Delta_n \mathbb{E}_{\theta_0}\left(g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\\ &\qquad+ \Delta_n \mathbb{E}_{\theta_0}\left((X_{t_i^n} - X_{t_{i-1}^n}) R(X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right)\\ &\qquad+ \Delta^2 \mathbb{E}_{\theta_0}\left(R(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)(W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n}\right).\end{aligned} \tag{5.18}$$

Note that

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

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

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

for suitable constants $C > 0$, with

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

for $m = 4, 5$ by Lemma A.3. Now, by (5.18), it only remains to show that

$$\frac{1}{\sqrt{\Delta_n}} \sum_{i=1}^{[nt]} \partial_y^2 g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta_0) \mathbb{E}_{\theta_0} \left((X_{t_i^n} - X_{t_{i-1}^n})^2 (W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) \xrightarrow{\mathcal{P}} 0 . \tag{5.19}$$

Applying Itô's formula with the function

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

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

$$\begin{aligned}
&(X_{t_i^n} - X_{t_{i-1}^n})^2 (W_{t_i^n} - W_{t_{i-1}^n}) \\
&\quad = 2 \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n})(W_s - W_{t_{i-1}^n}) a(X_s) \, ds + \int_{t_{i-1}^n}^{t_i^n} (W_s - W_{t_{i-1}^n}) b^2(X_s; \theta_0) \, ds \\
&\qquad + 2 \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n}) b(X_s; \theta_0) \, ds + 2 \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n})(W_s - W_{t_{i-1}^n}) b(X_s; \theta_0) \, dW_s \\
&\qquad + \int_{t_{i-1}^n}^{t_i^n} (X_s - X_{t_{i-1}^n})^2 \, dW_s .
\end{aligned} \tag{5.20}$$

By the martingale property of the Itô integrals in (5.20),

$$\begin{aligned}
&\mathbb{E}_{\theta_0} \left((X_{t_i^n} - X_{t_{i-1}^n})^2 (W_{t_i^n} - W_{t_{i-1}^n}) \mid \mathcal{F}_{t_{i-1}^n} \right) \\
&\quad = 2 \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((X_s - X_{t_{i-1}^n})(W_s - W_{t_{i-1}^n}) a(X_s) \mid \mathcal{F}_{t_{i-1}^n} \right) ds \\
&\qquad + \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((W_s - W_{t_{i-1}^n}) b^2(X_s; \theta_0) \mid \mathcal{F}_{t_{i-1}^n} \right) ds \\
&\qquad + 2 \int_{t_{i-1}^n}^{t_i^n} \mathbb{E}_{\theta_0} \left((X_s - X_{t_{i-1}^n}) b(X_s; \theta_0) \mid X_{t_{i-1}^n} \right) ds .
\end{aligned} \tag{5.21}$$

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

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

and by Lemma 2.6

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

so also

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

Now

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

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

by Lemma A.3, so by (5.19) and (5.21), it remains to show that

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

Applying Itô's formula with the function

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

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

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

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

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

Now

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

thus completing the proof. □

Proof of Lemma 5.3. The aim of this proof is to establish that the conditions of Theorem IX.7.28 in Jacod and Shiryaev (2003) hold, by which the desired result follows directly.

For all $t>0$,

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

and since the right-hand side converges to 0 in probability under $\mathbb{P}_{\theta_0}$ by (5.1) of Lemma 5.1, so does the left-hand side, i.e. condition 7.27 of Theorem IX.7.28 holds. From (5.2) and (5.4) of Lemma 5.1, it follows that for all $t > 0$,

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

establishing that condition 7.28 of Theorem IX.7.28 is satisfied. Lemma 5.2 implies condition 7.29, while the Lyapunov condition (5.3) of Lemma 5.1 implies the Lindeberg condition 7.30 of Theorem IX.7.28 in Jacod and Shiryaev (2003), from which the desired result now follows.

Theorem IX.7.28 contains an additional condition 7.31. This condition has the same form as (5.5), but with $W_{t_i^n} - W_{t_{i-1}^n}$ replaced by $N_{t_i^n} - N_{t_{i-1}^n}$, where $\mathbf{N} = (N_t)_{t\geq 0}$ is any bounded martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_{\theta_0})$, which is orthogonal to $\mathbf{W}$. However, since $(\mathcal{F}_t)_{t\geq 0}$ is generated by U and $\mathbf{W}$, it follows from the martingale representation theorem (Jacod and Shiryaev, 2003, Theorem III.4.33) that every martingale on $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P}_{\theta_0})$ may be written as the sum of a constant term and a stochastic integral with respect to $\mathbf{W}$, and therefore cannot be orthogonal to $\mathbf{W}$. □

A Auxiliary Results

This section contains a number of technical results used in the proofs in Section 5.2.

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

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

for some random variable F, then

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

◇

Lemma A.2. *Suppose that Assumptions 2.4 and 2.5 hold. Then, for all $\theta \in \Theta$,*

(i)

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

(ii)

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

(iii)

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

(iv)

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

(v)

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

$\diamond$

Proof of Lemma A.2. The formulae (A.1), (A.2) and (A.3) are implicitly given in the proofs of Sørensen (2010, Lemmas 3.2 & 3.4). To prove the two remaining formulae, note first that using (2.5), Assumption 2.5.(i) and Lemma 2.7,

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

The verification of these formulae may be simplified by using e.g. the Leibniz formula for the n'th derivative of a product to see that partial derivatives are zero when evaluated in $y = x$. These results, as well as Lemmas 2.6 and 2.7, and (A.8) are used without reference in the following.

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

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

proving (A.4). Similarly,

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

which proves (A.5). □

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

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

◇

Lemma A.3 follows easily by the convergence of Riemann sums.

Lemma A.4. *Suppose that Assumption 2.4 holds, and let $m \geq 2$. Then, there exists a constant $C_m > 0$, such that for $0 \leq t \leq t + \Delta \leq 1$,*

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

$\diamond$

Corollary A.5. *Suppose that Assumption 2.4 holds. Let a compact, convex set $K \subseteq \Theta$ be given, and suppose that $f(y, x; \theta)$ is of polynomial growth in x and y, uniformly for θ in K. Then, there exist constants $C_K > 0$ such that for $0 \leq t \leq t + \Delta \leq 1$,*

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

for all $\theta \in K$. $\diamond$

Lemma A.4 and Corollary A.5, correspond to Lemma 6 of Kessler (1997), adapted to the present assumptions. For use in the following, observe that for any $\theta \in \Theta$, there exist constants $C_\theta > 0$ such that

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

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

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

Note that by Corollary A.5, it holds that under Assumption 2.4,

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

Lemma A.6. *Suppose that Assumption 2.4 holds, and that the function $f(t, y, x; \theta)$ satisfies that*

$$f(t, y, x; \theta) \in C^{pol}_{1,2,1}([0,1] \times \mathcal{X}^2 \times \Theta) \quad \textit{with} \quad f(0, x, x; \theta) = 0 \tag{A.9}$$

for all $x \in \mathcal{X}$ and $\theta \in \Theta$. Let $m \in \mathbb{N}$ be given, and let $Dk(\,\cdot\,; \theta, \theta') = k(\,\cdot\,; \theta) - k(\,\cdot\,; \theta')$. Then, there exist constants $C_m > 0$ such that

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

for $0 \leq s < t \leq 1$ and $\theta, \theta' \in \Theta$, where f_1 and f_2 are given by

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

$$f_2(t,y,x;\theta) = b(y;\theta_0)\partial_y f(t,y,x;\theta)\ .$$

Furthermore, for each compact, convex set $K \subseteq \Theta$, there exists a constant $C_{m,K} > 0$ such that

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

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

Proof of Lemma A.6. A simple application of Itô's formula (when conditioning on $X_s = x_s$) yields that for all $\theta \in \Theta$,

$$f(t-s,X_t,X_s;\theta) = \int_s^t f_1\,(u-s,X_u,X_s;\theta)\,du + \int_s^t f_2\,(u-s,X_u,X_s;\theta)\,dW_u \tag{A.11}$$

under $\mathbb{P}_{\theta_0}$.

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

$$\mathbb{E}_{\theta_0}\left(\left|\int_s^t Df_j(u-s,X_u,X_s;\theta,\theta')^j\,du\right|^k\right) \leq (t-s)^{k-1}\int_s^t \mathbb{E}_{\theta_0}\left(\left|Df_j(u-s,X_u,X_s;\theta,\theta')\right|^{jk}\right)du \tag{A.12}$$

for $j=1,2$, and by the martingale properties of the second term in (A.11), the Burkholder-Davis-Gundy inequality may be used to show that

$$\mathbb{E}_{\theta_0}\left(\left|\int_s^t Df_2(u-s,X_u,X_s;\theta,\theta')\,dW_u\right|^{2m}\right) \leq C_m\mathbb{E}_{\theta_0}\left(\left|\int_s^t Df_2(u-s,X_u,X_s;\theta,\theta')^2\,du\right|^m\right). \tag{A.13}$$

Now, (A.11), (A.12) and (A.13) may be combined to show (A.10). The last result of the lemma follows by an application of the mean value theorem. $\square$

Lemma A.7. *Suppose that Assumption 2.4 holds, and let $K \subseteq \Theta$ be compact and convex. Assume that $f(t,y,x;\theta)$ satisfies (A.9) for all $x \in \mathcal{X}$ and $\theta \in \Theta$, and define*

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

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

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

for all $\theta,\theta' \in K$ and $n \in \mathbb{N}$. Define $\widetilde{F}_n(\theta) = \Delta_n^{-1}F_n(\theta)$, and suppose, moreover, that the functions

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

satisfy (A.9) for $j = 1, 2$. Then, for each $m \in \mathbb{N}$, there exists a constant $C_{m,K} > 0$, such that

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

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

Proof of Lemma A.7. For use in the following, define, in addition to h_1, h_2 and h_{j2}, the functions

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

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

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

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

First, by the martingale properties of

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

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

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

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

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

thus

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

for all $\theta, \theta' \in K$ and $n \in \mathbb{N}$. In the case where also h_j and h_{j2} satisfy (A.9) for all $x \in \mathcal{X}$, $\theta \in \Theta$ and $j = 1, 2$, use Lemma A.6 to write

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

and similarly obtain

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

Now, inserting into (A.14),

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

and, ultimately,

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

□

Lemma A.8. *Suppose that Assumption 2.4 is satisfied. Let $f \in C^{pol}_{0,1}(\mathcal{X} \times \Theta)$. Define*

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

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

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

$\diamond$

Lemma A.8 follows from a simple application of the mean value theorem.

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

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

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

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

Then,

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

$\diamond$

Proof of Lemma A.9. $(H_n(\theta))_{n\in\mathbb{N}}$ is tight in $\mathbb{R}$ for all $\theta \in K$, so, using (A.15), it follows from Kallenberg (2002, Corollary 16.9 & Theorem 16.3) that the sequence of processes $(\mathbf{H}_n)_{n\in\mathbb{N}}$ is tight in $C(K,\mathbb{R})$, the space of continuous (and bounded) real-valued functions on K, and thus relatively compact in distribution. Also, for all $d \in \mathbb{N}$ and $(\theta_1,\ldots,\theta_d) \in K^d$,

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

so by Kallenberg (2002, Lemma 16.2), $\mathbf{H}_n \xrightarrow{\mathcal{D}} 0$ in $C(K,\mathbb{R})$ equipped with the uniform metric. Finally, by the continuous mapping theorem, $\sup_{\theta\in K}|H_n(\theta)| \xrightarrow{\mathcal{D}} 0$, and the desired result follows. □

Acknowledgement

Nina Munkholt Jakobsen was supported by the Danish Council for Independent Research | Natural Science through a grant to Susanne Ditlevsen. Michael Sørensen was supported by the Center for Research in Econometric Analysis of Time Series funded by the Danish National Research Foundation. The research is part of the Dynamical Systems Interdisciplinary Network funded by the University of Copenhagen Programme of Excellence.

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. (2008). Closed form likelihood expansions for multivariate diffusions. *Annals of Statistics*, 36(2):906–937.

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.

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

Bibbona, E., Lansky, P., and Sirovich, R. (2010). Estimating input parameters from intracellular recordings in the Feller neuronal model. *Physical Review E*, 81(3):031916.

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

Bladt, M., Finch, S., and Sørensen, M. (2015). Simulation of multivariate diffusion bridges. To appear in J. Roy. Statist. Soc. B.

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

Cox, J. C., Ingersoll, Jr., J. E., and Ross, S. A. (1985). A theory of the term structure of interest rates. *Econometrica*, 53(2):385–407.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19(4):263–284.

Ditlevsen, S. and Lansky, P. (2006). Estimation of the input parameters in the Feller neuronal model. *Physical Review E*, 73(6):061910.

Dohnal, G. (1987). On estimating the diffusion coefficient. *Journal of Applied Probability*, 24(1):105–114.

Durham, G. B. and Gallant, A. R. (2002). Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes. *Journal of Business & Economic Statistics*, 20(3):297–316.

Favetto, B. and Samson, A. (2010). Parameter estimation for a bidimensional partially observed Ornstein-Uhlenbeck process with biological application. *Scandinavian Journal of Statistics*, 37(2):200–220.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, 20(4):547–557.

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

Forman, J. L. and Sørensen, M. (2008). The Pearson diffusions: A class of statistically tractable diffusion processes. *Scandinavian Journal of Statistics*, 35(3):438–465.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multi-dimensional diffusion processes. *Annales de l'institut Henri Poincaré (B) Probabilités et Statistiques*, 29(1):119–151.

Genon-Catalot, V. and Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: Random sampling. *Scandinavian Journal of Statistics*, 21(3):193–221.

Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic Processes and their Applications*, 119(3):679–699.

Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. *Bernoulli*, 7(6):899–912.

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

Golightly, A. and Wilkinson, D. J. (2006). Bayesian sequential inference for nonlinear multivariate diffusions. *Statistics and Computing*, 16:323–338.

Golightly, A. and Wilkinson, D. J. (2008). Bayesian inference for nonlinear multivariate diffusion models observed with error. *Computational Statistics and Data Ananysis*, 52:1674–1693.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application*. New York, Springer.

Iacus, S. M. (2014). *sde: Simulation and Inference for Stochastic Differential Equations*. R package version 2.0.13. http://CRAN.R-project.org/package=sde.

Jacobsen, M. (2001). Discretely observed diffusions: Classes of estimating functions and small Delta-optimality. *Scandinavian Journal of Statistics*, 28(1):123–149.

Jacobsen, M. (2002). Optimality and small Delta-optimality of martingale estimating functions. *Bernoulli*, 8(5):643–668.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. In *Seminaire de Probabilites XXXI*, pages 232–246. Heidelberg, Springer.

Jacod, J. (2000). Non-parametric kernel estimation of the coefficient of a diffusion. *Scandinavian Journal of Statistics*, 27(1):83–96.

Jacod, J. (2006). Parametric inference for discretely observed non-ergodic diffusions. *Bernoulli*, 12(3):383–401.

Jacod, J. (2010). Inference for stochastic processes. In Aït-Sahalia, Y. and Hansen, L. P., editors, *Handbook of Financial Econometrics: Applications*, volume 2 of *Handbooks in Finance*, chapter 16, pages 197–239. Amsterdam, Elsevier.

Jacod, J. and Protter, P. (1998). Asymptotic error distributions for the Euler method for stochastic differential equations. *Annals of Probability*, 26(1):267–307.

Jacod, J. and Shiryaev, A. (2003). *Limit Theorems for Stochastic Processes, 2nd edition*. New York, Springer.

Jacod, J. and Sørensen, M. (2012). Aspects of asymptotic statistical theory for stochastic processes. Preprint, University of Copenhagen, Denmark.

Jeganathan, P. (1982). On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā: The Indian Journal of Statistics, Series A*, pages 173–212.

Kallenberg, O. (2002). *Foundations of Modern Probability, 2nd Edition*. New York, Springer.

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

Kessler, M., Lindner, A., and Sørensen, M. (2012). *Statistical Methods for Stochastic Differential Equations*. Boca Raton, Chapman & Hall.

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

Larsen, K. S. and Sørensen, M. (2007). A diffusion model for exchange rates in a target zone. *Mathematical Finance*, 17:285–306.

Le Cam, L. and Yang, G. L. (2000). *Asymptotics in Statistics: Some Basic Concepts, 2nd edition*. New York, Springer.

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

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413.

Mykland, P. A. and Zhang, L. (2006). ANOVA for diffusions and Ito processes. *Annals of Statistics*, 34(4):1931–1963.

Pedersen, A. R. (1995). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71.

Pedersen, A. R. (2000). Estimating the nitrous oxide emission rate from the soil surface by means of a diffusion model. *Scandinavian Journal of Statistics*, 27(3):385–403.

Picchini, U., Ditlevsen, S., and De Gaetano, A. (2008). Maximum likelihood estimation of a time-inhomogeneous stochastic differential model of glucose dynamics. *Mathematical Medicine and Biology*, 25(2):141–155.

Prakasa Rao, B. L. S. (1983). Asymptotic theory for non-linear least squares estimator for diffusion processes. *Statistics*, 14(2):195–209.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

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

Sørensen, H. (2004). Parametric inference for diffusion processes observed at discrete points in time: A survey. *International Statistical Review*, 72:337 – 354.

Sørensen, M. (2010). Efficient estimation for ergodic diffusions sampled at high frequency. Preprint, University of Copenhagen, Denmark.

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

Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scandinavian Journal of Statistics*, 31(4):553–566.

Uchida, M. and Yoshida, N. (2013). Quasi likelihood analysis of volatility and nondegeneracy of statistical random fields. *Stochastic Processes and Their Applications*, 123:2851–2876.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, 5(2):177–188.

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