Efficient estimation for ergodic diffusions sampled at high frequency

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

Abstract

A general theory of efficient estimation for ergodic diffusions sampled at high frequency is presented. High frequency sampling is now possible in many applications, in particular in finance. The theory is formulated in term of approximate martingale estimating functions and covers a large class of estimators including most of the previously proposed estimators for diffusion processes, for instance GMM-estimators and the maximum likelihood estimator. Simple conditions are given that ensure efficiency and rate optimality, where estimators of parameters in the diffusion coefficient take advantage of the information in the quadratic variation and therefore converge faster than estimators of parameters in the drift coefficient. These conditions facilitates the choice of estimators from the confusing multitude of estimators that have been proposed for diffusion models. The conditions turn out to be equal to those implying small Δ-optimality in the sense of Jacobsen. Thus the theory presented here provides an interpretation of Jacobsen's concept in terms of classical statistical concepts. Optimal martingale estimating functions in the sense of Godambe and Heyde are, under weak conditions, shown to give rate optimal and efficient estimators.

Key words: Approximate martingale estimating functions, discrete time observation of a diffusion, efficiency, Euler approximation, generalized method of moments, optimal estimating function, optimal rate, small delta-optimality.

1 Introduction

Dynamic phenomena affected by random noise are often modelled in continuous time by stochastic differential equations. Among the advantages of this approach are model parameters with a clear interpretation and facilitation of communication with engineers and scientists by a common modelling tool, differential equations. Finance is a well-known example of an area where stochastic differential equations are widely used. A few other examples are agronomy (Pedersen (2000)), climatology (Ditlevsen, Ditlevsen & Andersen (2002)), gene regulation (McAdams & Arkin (1997)), molecular dynamics (Pokern, Stuart & Wiberg (2009)), neurology (Lansky, Sacerdote & Tomasetti (1995)), and physiology (Ditlevsen et al. (2007)). While the dynamics is formulated in continuous time, observations are at discrete points in time. Statistical inference for these models has in recent years become an intensive area of research, and a profusion of estimators have been proposed for parametric diffusion models, see e.g. Sørensen (2004). A large number of simulation studies have been performed to compare the relative merits of various estimators, but the general picture has so far remained rather confusing. In the present paper, simple and easily checked criteria are derived for efficiency and rate optimality of estimators. The latter property is crucial for diffusion models, as will be explained below. The criteria derived here provide, in combination with considerations of computing time, much needed clarity in this area of statistics.

Our focus is on approximate martingale estimating functions for discrete time observations of a diffusion process. This approach covers most of the previously proposed estimators, and the few that are not covered are likely to be less efficient, because the non-martingale estimating functions, in general, do not approximate the score function as well as martingales. We consider a scalar diffusion given by the stochastic differential equation

$$dX_t = b(X_t; \alpha)dt + \sigma(X_t; \beta)dW_t, \tag{1.1}$$

where $(\alpha, \beta) = \theta \in \Theta \subseteq \mathbb{R}^2$ are parameters to be estimated. The restriction to a scalar process and two scalar parameters is done to simplify the presentation. The results can be generalized to multivariate parameters and diffusions. In Section 3 we indicate how results for multivariate diffusions differ from the one-dimensional case. Martingale estimating functions give consistent estimators at all sampling frequencies, Bibby, Jacobsen & Sørensen (2010), and Godambe-Heyde optimal martingale estimating functions have turned out to often provide estimators with a high efficiency, see e.g. Overbeck & Rydén (1997) and Larsen & Sørensen (2007). One aim of this paper is to explain this by showing that the estimators are efficient in a high frequency asymptotic scenario. In particular, it is shown that optimal martingale estimating functions in the sense of Godambe & Heyde (1987) give rate optimal and efficient estimators under weak regularity conditions. Thus they have an unexpected global optimality property.

The observation times are $i\Delta_n$, $i = 0, \dots, n$ and the asymptotic scenario considered is that

$$n \to \infty, \qquad \Delta_n \to 0, \qquad n\Delta_n \to \infty.$$

The length of the time interval in which observations are made goes to infinity, which is necessary to ensure that the drift parameter α can be estimated consistently. If the drift coefficient is known, this condition is not needed. At the same time the sampling frequency goes to infinity, which allows us to study how the special structure of diffusion models implies that martingale estimating functions can yield efficient estimators. For estimating functions that are not exact martingales, we need the extra condition that Δ_n goes to zero sufficiently fast to ensure that $n\Delta^\kappa \to 0$, for a certain κ that depends on how far the estimating function is from being a martingale. This allows us to apply the central limit theorem for martingales. That our high frequency asymptotics is relevant to applications is due to the fact that the sampling frequency needs not be particularly high for the asymptotics to be applicable, provided that the diffusion does not move too fast. This is, for instance, often the case for finance data, where even weekly observations can in some cases be considered a high sampling frequency. This explains why estimators from optimal martingale estimating functions have quite often been found to have good efficiency in finance applications; see e.g. Larsen & Sørensen (2007).

We consider estimating functions of the general form

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_n, X_{i\Delta_n}, X_{(i-1)\Delta_n}; \theta), \tag{1.2}$$

where the function $g(\Delta, y, x; \theta)$ with values in $\mathbb{R}^2$ is such that G_n is, exactly or approximately, a martingale estimating function. Thus theory developed in this paper covers a large class of estimators including most of the previously proposed estimators for diffusion processes, for instance the martingale estimating functions proposed by Bibby & Sørensen (1995) and Kessler & Sørensen (1999), GMM-estimators based on conditional moments, Hansen (1982), Hansen (1985) and Hansen (1993), and the maximum likelihood estimator, Pedersen (1995), Poulsen (1999), Aït-Sahalia (2002), Durham & Gallant (2002), Aït-Sahalia & Mykland (2003) and Beskos et al. (2006). The pseudo-likelihood function obtained from the Gaussian Euler approximation to the transition density is covered too. This pseudo-likelihood can, when β is fixed, also be obtained as a discretization of the continuous time likelihood function. These estimators have often been used in empirical work in finance. Estimators closely related to the Euler pseudo-likelihood were considered by Prakasa Rao (1988), Florens-Zmirou (1989), and Yoshida (1992). Also more complex pseudo-likelihood functions are covered such as those proposed by Kessler (1997), who obtained more accurate Gaussian approximations to the likelihood function by higher order expansions of conditional moments. The latter group of authors considered the same high frequency asymptotic scenario as the one in the present paper. Sørensen & Uchida (2003) and Gloter & Sørensen (2009) considered the Euler pseudo likelihood under a combination of high frequency and small diffusion asymptotics, where the diffusion coefficient goes to zero as $n \to \infty$. Under this asymptotic scenario the infinite horizon condition, $n\Delta_n \to \infty$, is not needed for consistent estimation of α.

We find very simple conditions on the function $g(\Delta, y, x; \theta)$ that ensure rate optimality and efficiency of estimators. The condition for *rate optimality* is

Condition 1.1

$$\partial_y g_2(0, x, x; \theta) = 0 \tag{1.3}$$

for all x in the state-space of the diffusion process and all $\theta \in \Theta$.

By $\partial_y g_2(0, x, x; \theta)$ we mean $\partial_y g_2(0, y, x; \theta)$ evaluated at $y = x$. For diffusion models rate optimality is important because the diffusion coefficient parameter β can be estimated at a higher rate than the drift parameter α; see Gobet (2002). Estimators of β that do not use the information about the diffusion coefficient contained in the quadratic variation will converge at the same, relatively slow, rate as estimators of α. We will refer to (1.3) as *Jacobsen's condition* because it equals one of the conditions for small Δ-optimality obtained in Jacobsen (2002) for martingale estimating functions. The idea behind the small Δ-optimality concept, introduced in Jacobsen (2001), is to consider the asymptotic covariance matrix obtained under a low frequency asymptotics, where the time between observations, Δ, is fixed and does not depend on n. This asymptotic covariance matrix is expanded in powers of Δ, and the small Δ-optimal estimating functions are those for which the leading term of the expansion is minimized within the class of all estimating functions. The same kind of reasoning was used by Aït-Sahalia & Mykland (2004) to study observation at random time points. In Jacobsen's approach the Condition 1.1 was introduced to avoid a singularity in the asymptotic variance when the time between observations tends to zero. In our high frequency approach, the condition implies rate optimality for estimation of the diffusion coefficient parameter and can also be interpreted as a condition that ensures that the estimating function can estimate quadratic variation.

The condition for *efficiency* obtained in this paper is

Condition 1.2

$$\partial_y g_1(0, x, x; \theta) = \partial_\alpha b(x; \alpha)/\sigma^2(x; \beta) \tag{1.4}$$

and

$$\partial_y^2 g_2(0, x, x; \theta) = \partial_\beta \sigma^2(x; \beta)/\sigma^4(x; \beta), \tag{1.5}$$

for all x in the state space of the diffusion process and all $\theta \in \Theta$.

Also (1.4) and (1.5) were found as conditions for small Δ-optimality in Jacobsen (2002). This is not surprising. Our results provide an interpretation of small Δ-optimality in terms of the classical statistical concepts rate optimality and efficiency.

The paper is organized as follows. Section 2 sets up the model, the class of approximate martingale estimating functions, and the assumptions used throughout the paper. A number of often used estimators are shown to be covered by the theory. Also a crucial fundamental lemma is presented. Section 3 develops the high frequency asymptotic theory for general estimating functions and, more importantly, for rate optimal estimating functions. The asymptotic results are used in Section 4 to find conditions for efficiency. Sufficient conditions that a given set of conditional moments can give a rate optimal and efficient estimator are given, and it is proved that Godambe-Heyde optimal martingale estimating functions are rate optimal and efficient. A number of examples are considered, including the Euler pseudo-likelihood and maximum likelihood estimation. Proofs are given in Section 5, where tools of some independent interest for studying high frequency asymptotic properties of estimators are provided. Section 6 concludes.

2 Model and conditions

We consider observations $X_{t_0^n}, \ldots, X_{t_n^n}$ of the process given by (1.1) at the time points $t_i^n = i\Delta_n$, $i = 0, \ldots, n$. We suppose that a solution of the stochastic differential equation (1.1) exists, is unique in law, and is adapted to the filtration generated by the Wiener process W and the initial value X_0. To simplify the presentation, we assume that α and β are one-dimensional. All results in the paper can be immediately generalized to the case where α and β are multivariate by replacing partial derivatives by vectors or matrices of partial derivative and by considering estimating functions of the same dimension as the parameter. We assume further that $\theta = (\alpha, \beta) \in \Theta$ where Θ is a subset of $\mathbb{R}^2$ with a non-empty interior $\operatorname{int}\Theta$, and that the true parameter value $\theta_0 = (\alpha_0, \beta_0) \in \operatorname{int}\Theta$. It is no serious restriction to assume that Θ is convex. The theory and results involve the squared diffusion coefficient

$$v(x; \beta) = \sigma^2(x; \beta) \tag{2.1}$$

rather than the diffusion coefficient. We denote the state-space of X by (ℓ, r), where $-\infty \leq \ell < r \leq \infty$. We assume that $v(x; \beta) > 0$ for all $x \in (\ell, r)$, and that the stochastic differential equation (1.1) satisfies the following condition.

Condition 2.1 *The following holds for all $\theta \in \Theta$:*

(1)

$$\int_{x^{\#}}^{r} s(x;\theta)dx = \int_{\ell}^{x^{\#}} s(x;\theta)dx = \infty \tag{2.2}$$

and

$$\int_{\ell}^{r} x^k \tilde{\mu}_\theta(x)dx < \infty \tag{2.3}$$

for all $k \in \mathbb{N}$, where $x^{\#}$ is an arbitrary point in (ℓ, r),

$$s(x;\theta) = \exp\left(-2\int_{x^{\#}}^{x} \frac{b(y;\alpha)}{v(y;\beta)}dy\right) \tag{2.4}$$

and

$$\tilde{\mu}_\theta(x) = [s(x;\theta)v(x;\beta)]^{-1}. \tag{2.5}$$

(2) $\sup_t E_\theta(|X_t|^k) < \infty$ *for all $k \in \mathbb{N}$.*

(3) $b, \sigma \in C_{p,4,1}((\ell, r) \times \Theta)$.

(4) There exists a constant C_θ such that for all $x, y \in (\ell, r)$

$$|b(x;\alpha) - b(y;\alpha)| + |\sigma(x;\beta) - \sigma(y;\beta)| \leq C_\theta |x - y|$$

We define $C_{p,k_1,k_2,k_3}(\mathbb{R}_+ \times (\ell, r)^2 \times \Theta)$ as the class of real functions $f(t, y, x; \theta)$ satisfying that

(i) $f(t, y, x; \theta)$ is k_1 times continuously differentiable with respect t, k_2 times continuously differentiable with respect y, and k_3 times continuously differentiable with respect α and with respect to β

(ii) f and all partial derivatives $\partial_t^{i_1} \partial_y^{i_2} \partial_\alpha^{i_3} \partial_\beta^{i_4} f$, $i_j = 1, \ldots k_j$, $j = 1, 2$, $i_3 + i_4 \leq k_3$, are of polynomial growth in x and y uniformly for θ in a compact set (for fixed t).

The classes $C_{p,k_1,k_2}((\ell, r) \times \Theta)$ and $C_{p,k_1,k_2}((\ell, r)^2 \times \Theta)$ are defined similarly for functions $f(y;\theta)$ and $f(y, x;\theta)$, respectively. A function $f(y, x;\theta)$ is said to be of polynomial growth in y and x uniformly for θ in a compact set if, for any compact subset $K \subseteq \Theta$, there exists a constant $C > 0$ such that $\sup_{\theta \in K} |f(y, x;\theta)| \leq C(1 + |x|^C + |y|^C)$ for all x and y in the state-space of the diffusion.

The conditions (2.2) and (2.3) with $k = 1$ ensure that the process X is ergodic with invariant measure with Lebesgue density

$$\mu_\theta(x) = \tilde{\mu}_\theta(x) / \int_{\ell}^{r} \tilde{\mu}_\theta(y)dy. \tag{2.6}$$

Actually, (2.2) is not necessary. For instance if ℓ is finite and $\int_{\ell}^{x^{\#}} s(x;\theta) < \infty$, then the process can hit ℓ at a finite time with positive probability, but if the boundary is instantaneously reflecting, X is also ergodic in this case. To avoid worrying about making assumptions about the boundary behaviour, we impose the condition (2.2) under which the boundaries cannot be reached in finite time.

If $X_0 \sim \mu_\theta$, then the process is stationary and Condition 2.1 (2) follows trivially from (2.3). For diffusions with a spectral gap, which is frequently the case for models used in practice, Condition 2.1 (2) follows from (2.3), provided that $E_\theta(|X_0|^k) < \infty$. The solution to (1.1) is said to have a spectral gap if the smallest positive eigenvalue λ_θ of the generator

$$L_\theta = b(x;\alpha)\frac{d}{dx} + \frac{1}{2}v(x;\beta)\frac{d^2}{dx^2} \tag{2.7}$$

is strictly positive. Simple conditions ensuring this were given by Genon-Catalot, Jeantheau & Larédo (2000). It is, for instance, the case when the drift is linear, see e.g. Hansen, Scheinkman & Touzi (1998). With $\mu_k(\theta) = \int |x|^k \mu_\theta(x)dx$, it follows from the contraction property of the transition operator that

$$\int_A \left(E_\theta(|X_t|^k \mid X_0 = x) - \mu_k(\theta)\right)^2 \mu_\theta(x)dx \leq \int_A \left(|x|^k - \mu_k(\theta)\right)^2 \mu_\theta(x)dx,$$

for any Borel subset $A \subseteq (\ell, r)$. Hence $|E_\theta(|X_t|^k \mid X_0 = x) - \mu_k(\theta)| \leq ||x|^k - \mu_k(\theta)|$ for almost all $x \in (\ell, r)$, which shows that Condition 2.1 (2) is satisfied when $E_\theta(|X_0|^k) < \infty$.

We consider estimating functions of the general form (1.2) where the function $g(\Delta, y, x; \theta)$ with values in $\mathbb{R}^2$ satisfies the following condition.

Condition 2.2

(1) For a $\kappa \geq 2$

$$E_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \mid X_{t_{i-1}^n}) = \Delta_n^\kappa R(\Delta_n, X_{t_{i-1}^n}; \theta) \quad \textit{for all } \theta \in \Theta. \tag{2.8}$$

(2) The function $g(\Delta, y, x; \theta)$ *has an expansion in powers of* Δ

$$g(\Delta, y, x; \theta) = g(0, y, x; \theta) + \Delta g^{(1)}(y, x; \theta) + \tfrac{1}{2}\Delta^2 g^{(2)}(y, x; \theta) + \Delta^3 R(\Delta, y, x; \theta), \tag{2.9}$$

and

$$\begin{aligned}
g(\Delta, y, x; \theta) &\in C_{p,2,6,2}(\mathbb{R} \times (\ell, r)^2 \times \Theta),\\
g(0, y, x; \theta) &\in C_{p,6,2}((\ell, r)^2 \times \Theta)\\
g^{(1)}(y, x; \theta) &\in C_{p,4,2}((\ell, r)^2 \times \Theta),\\
g^{(2)}(y, x; \theta) &\in C_{p,2,2}((\ell, r)^2 \times \Theta).
\end{aligned}$$

Here and in the rest of the paper, $R(\Delta, y, x; \theta)$ denotes a (generic) function such that $|R(\Delta, y, x; \theta)| \leq F(y, x; \theta)$ where F is of polynomial growth in y and x uniformly for θ in a compact set. Similarly for $R(\Delta, x; \theta)$. The assumptions of polynomial growth are made only to simplify the presentation of the theory. These assumptions are satisfied for most models used in practice, but the results hold under weaker assumptions as long as the necessary moments exist and the remainder terms can be controlled so that we have expansions to the orders needed in the proofs. Presumably the condition can be weakened in a way similar to the method used in Gloter & Sørensen (2009).

We remind the reader of the trivial fact that for any non-singular 2×2 matrix, M_n, the estimating functions $M_n G_n(\theta)$ and $G_n(\theta)$ give exactly the same estimator. We call them *versions* of the same estimating function. The matrix M_n may depend on Δ_n. Therefore a given version of an estimating function needs not satisfy Condition 2.2. The point is that a version must exist that satisfies the condition. It may for instance be necessary to multiply one of the coordinates by Δ_n. Examples of this phenomenon will be given in Section 4. The same remark can be made about other conditions later in the paper. A version must exists that satisfies all necessary conditions for a given result.

We shall often apply the generator (2.7) to a function $h(y, x)$ of two variables. This will always be taken to mean the following

$$L_\theta(h)(y,x) = b(y;\alpha)\partial_y h(y,x) + \tfrac{1}{2}v(y;\beta)\partial_y^2 h(y,x). \tag{2.10}$$

For a function $h(\Delta, y, x; \theta)$ that depends also on Δ and θ, we use the notation

$$L_\theta(h(\Delta;\tilde{\theta}))(y,x) = b(y;\alpha)\partial_y h(\Delta,y,x;\tilde{\theta}) + \tfrac{1}{2}v(y;\beta)\partial_y^2 h(\Delta,y,x;\tilde{\theta}).$$

The following lemma provides identities that play an essential role in the proofs of the asymptotic theory in the next section. The identities are a consequence of the approximate martingale property (2.8).

Lemma 2.3 *Under the Conditions 2.1 and 2.2*

$$g(0,x,x;\theta) = 0 \tag{2.11}$$
$$g^{(1)}(x,x;\theta) = -L_\theta(g(0;\theta))(x,x) \tag{2.12}$$

for all $x \in (\ell, r)$ and $\theta \in \Theta$. If $\kappa \geq 3$,

$$g^{(2)}(x,x;\theta) = -L_\theta^2(g(0;\theta))(x,x) - 2L_\theta(g^{(1)}(\theta))(x,x). \tag{2.13}$$

The identity (2.13) is not used in the rest of the paper, but is useful if expansions of a higher order are needed.

2.1 Examples

A main example of estimating functions that satisfy condition (2.8) are the *martingale estimating functions* for which

$$E_\theta(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}) = 0.$$

They often have the form

$$g(\Delta,y,x;\theta) = \sum_{j=1}^{N} a_j(x,\Delta;\theta)[f_j(y;\theta) - E_\theta(f_j(X_\Delta;\theta) \mid X_0 = x)]. \tag{2.14}$$

A simple example is obtained for $N = 2$, $f_1(x) = x$ and $f_2(x) = x^2$. This *quadratic martingale estimating function* can be obtained as the pseudo score corresponding to a Gaussian approximate likelihood function, see Section 4. Other instances are polynomial estimating functions, where the functions f_j are power functions, and the estimating functions based on eigenfunctions of the generator proposed by Kessler & Sørensen (1999). With the specification (2.14) of g, the Condition 2.2 (2) is automatically satisfied provided the functions $f_j(x;\theta)$ are 6 times continuous differentiable with respect to x. To see this we need the result that for any $2(k+1)$ times differentiable function f

$$\begin{aligned} &\mathrm{E}_\theta(f(X_{t+s}) \mid X_t) \\ &\quad = \sum_{i=0}^{k} \frac{s^i}{i!} L_\theta^i f(X_t) + \int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_k} E_\theta(L_\theta^{k+1} f(X_{t+u_{k+1}}) \mid X_t) du_{k+1} \cdots du_1, \end{aligned} \tag{2.15}$$

where L_θ denotes the generator (2.7), see Florens-Zmirou (1989). Here we take the domain of L_θ to be the set of all twice continuously differentiable functions defined on the state space. That the conditional expectation in the remainder term is finite and that the remainder term has the right order follows from Lemma 5.1 in Section 5. Usually the weight functions a_j in (2.14) depend on Δ and must also be expanded to establish Condition 2.2 (2). For the specification (2.14), the conclusions of Lemma 2.3 trivially hold because in this case $g(0,y,x,\theta) = \sum_{j=1}^N a_j(x,0;\theta)[f_j(y) - f_j(x)]$, $g^{(1)}(x,x;\theta) = -\sum_{j=1}^N a_j(x,0;\theta) L_\theta f_j(x)$, and $g^{(2)}(x,x;\theta) = -\sum_{j=1}^N [a_j(x,0;\theta) L_\theta^2 f_j(x) + 2\partial_\Delta a_j(x,0;\theta) L_\theta f_j(x)]$.

The econometric *generalized method of moments* (GMM, see Hansen (1982)) based on conditional moments is covered by our theory. This method is in practice often implemented as follows; see Campbell, Lo & MacKinlay (1997). The starting point is an N-dimensional function $h(\Delta,y,x;\theta)$ each coordinate of which satisfies that $E_\theta(h_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}) = 0$. Let A_n be an $N \times N$-matrix such that $m_n(\theta) = A_n \sum_{i=1}^n h(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)$ converges in probability. For the usual low frequency asymptotics, where Δ_n does not depend on n, $A_n = n^{-1} I$, but for the high frequency asymptotics considered here, a different choice of A_n is usually necessary, as will be clear from the discussion in the next section. The GMM estimator is obtained by minimizing $Q_n(\theta) = m_n(\theta)^T W_n \, m_n(\theta)$, where W_n is an $N \times N$-matrix such that $W_n \to W$ in probability. Here and later x^T denotes the transpose of a vector or matrix x. The matrix W_n is typically the inverse of a consistent estimator of the covariance matrix of $m_n(\theta)$ (suitably normalized). Under weak regularity conditions, the GMM estimator solves the estimating equation $\partial_\theta Q_n(\theta) = \partial_\theta m_n(\theta)^T W_n \, m_n(\theta) = 0$, so if $\partial_\theta m_n(\theta) \to D(\theta)$ in probability (which is a necessary condition for asymptotic results about the GMM estimator), then the GMM estimator has the same asymptotic distribution as the estimator obtained from the martingale estimating function with

$$g(\Delta,y,x;\theta) = D(\theta)^T W h(\Delta,y,x;\theta).$$

This function will very often be of the form (2.14). The close relationship between martingale estimating functions and the type of GMM-estimators described here is discussed in detail in Christensen & Sørensen (2008). More general GMM-estimators of the martingale estimating function type were considered in Hansen (1985) and Hansen (1993).

Other estimating functions are obtained by replacing the exact conditional expectation in (2.14) by the expansion $\sum_{i=0}^{\kappa-1} s^i/i! \, L_\theta^i f(X_t)$. In this way a class of estimating functions is obtained that satisfies (2.8). Estimators obtained from this class include the simple example $g(\Delta,y,x;\theta) = a(x,\Delta;\theta)(y - b(x;\alpha)\Delta)$ with $\kappa = 2$ considered by Prakasa Rao (1988) and Florens-Zmirou (1989), the pseudo maximum likelihood estimators obtained from the

Gaussian Euler approximation to the likelihood, but also for instance, the estimators proposed by Chan et al. (1992) and Kelly, Platen & Sørensen (2004). For all $\kappa \in \mathbb{N}$, ($\kappa \geq 2$), Kessler (1997) proposed a *Gaussian approximation* to the likelihood function, for which the corresponding pseudo-score function is an approximate martingale estimating function that satisfies (2.8).

3 Optimal rate

In this section we give asymptotic results for approximate martingale estimating functions. It turns out that a very simple condition is needed to ensure rate optimal estimators, i.e. that the estimator of the parameter in the diffusion coefficient converges faster than the estimator of the parameter in the drift coefficient. As previously, x^T denotes the transpose of a vector or matrix x. We begin with a general approximate martingale estimating function.

Theorem 3.1 *Assume that the Conditions 2.1 and 2.2 hold. Suppose, moreover, the identifiability condition that*

$$\gamma(\theta,\theta_0) = \int_\ell^r [b(x,\alpha_0) - b(x,\alpha)]\partial_y g(0,x,x;\theta)\mu_{\theta_0}(x)dx \\ + \tfrac{1}{2}\int_\ell^r [v(x,\beta_0) - v(x,\beta)]\partial_y^2 g(0,x,x;\theta)\mu_{\theta_0}(x)dx \neq 0 \tag{3.1}$$

for all $\theta \neq \theta_0$, and that the matrix

$$S = \int_\ell^r A_{\theta_0}(x)\mu_{\theta_0}(x)dx \tag{3.2}$$

is invertible, where

$$A_\theta(x) = \begin{pmatrix} \partial_\alpha b(x;\alpha)\partial_y g_1(0,x,x;\theta) & \frac{1}{2}\partial_\beta v(x;\beta)\partial_y^2 g_1(0,x,x;\theta) \\ \partial_\alpha b(x;\alpha)\partial_y g_2(0,x,x;\theta) & \frac{1}{2}\partial_\beta v(x;\beta)\partial_y^2 g_2(0,x,x;\theta) \end{pmatrix} \tag{3.3}$$

Then a consistent estimator $\hat\theta_n = (\hat\alpha_n, \hat\beta_n)$ that solves the estimating equation $G_n(\theta) = 0$ exists and is unique in any compact subset of Θ containing θ_0 with a probability that goes to one as $n \to \infty$. For a martingale estimating function or more generally if $n\Delta^{2\kappa-1} \to 0$,

$$\sqrt{n\Delta_n}(\hat\theta_n - \theta_0) \xrightarrow{\mathcal{D}} N_2\left(0, S^{-1}V_0(S^T)^{-1}\right) \tag{3.4}$$

under P_{θ_0}, where $V_0 = V(\theta_0)$ with

$$V(\theta) = \int_\ell^r v(x,\beta_0)\partial_y g(0,x,x;\theta)\partial_y g(0,x,x;\theta)^T\mu_{\theta_0}(x)dx.$$

The theorem follows from the following lemma by asymptotic statistical results for stochastic processes, see e.g. Jacod & Sørensen (2010).

Lemma 3.2 *Under the Conditions 2.1 and 2.2*

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \xrightarrow{P_{\theta_0}} \gamma(\theta,\theta_0), \tag{3.5}$$

$$\begin{aligned}\frac{1}{n\Delta_n}\sum_{i=1}^{n} \partial_{\theta^T} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \quad &\xrightarrow{P_{\theta_0}} \\ &\int_\ell^r [L_{\theta_0}(\partial_\theta g(0;\theta))(x,x) - L_\theta(\partial_\theta g(0;\theta))(x,x) - A_\theta(x)]\mu_{\theta_0}(x)dx,\end{aligned} \tag{3.6}$$

and

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^T \xrightarrow{P_{\theta_0}} V(\theta), \tag{3.7}$$

uniformly when θ is in a compact set. For a martingale estimating function or more generally if $n\Delta^{2\kappa-1}\to 0$,

$$\frac{1}{\sqrt{n\Delta_n}}\sum_{i=1}^{n} g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \xrightarrow{\mathcal{D}} N_2\left(0, V_0\right). \tag{3.8}$$

Note that consistent estimators of $-S_0$ and V_0, and hence of the asymptotic variance of $\hat{\theta}_n$, can be obtained by inserting $\hat{\theta}_n$ into the left hand side of (3.6) and (3.7).

We see from (3.4) that the rate of convergence of both $\hat{\alpha}$ and $\hat{\beta}$ is $\sqrt{n\Delta_n}$, the square root of the length of the interval in which the diffusion is observed, when the matrix V_0 is regular. Gobet (2002) showed that under weak regularity conditions a discretely sampled diffusion model is local asymptotically normal in the high frequency asymptotic scenario considered here, and that the optimal rate of convergence for estimators of parameters in the drift coefficient is indeed $\sqrt{n\Delta_n}$, whereas the optimal rate for estimators of parameters in the diffusion coefficient is $\sqrt{n}$.

The next theorem shows that *Jacobsen's condition*, Condition 1.1, ensures rate optimal estimators. The reader is reminded that different versions of the estimating function give the same estimator, but will obviously not all satisfy Condition 1.1. The point is that for a given $g(\Delta, y, x;\theta)$ there must exist a version of the estimating function that satisfies the condition, i.e. there must exist a non-singular 2×2-matrix M, which may depend on Δ and θ, such that the second coordinate of $Mg(\Delta, y, x;\theta)$ satisfies (1.3). We will, for simplicity of presentation, assume that we start with a version that satisfies the condition. Similar remarks can be made about the conditions in the following theorem. The same version must satisfy all conditions.

Theorem 3.3 *Suppose the Conditions 2.1 and 2.2 hold, and that the second coordinate of g satisfies Jacobsen's Condition 1.1. Assume, moreover, that the following identifiability condition is satisfied*

$$\begin{aligned}\int_\ell^r [b(x,\alpha_0) - b(x,\alpha)]\partial_y g_1(0,x,x;\theta)\mu_{\theta_0}(x)dx \quad &\neq \quad 0 \qquad \textit{when } \alpha\neq\alpha_0 \\ \int_\ell^r [v(x,\beta_0) - v(x,\beta)]\partial_y^2 g_2(0,x,x;\theta)\mu_{\theta_0}(x)dx \quad &\neq \quad 0 \qquad \textit{when } \beta\neq\beta_0,\end{aligned}$$

and that $S_{11} \neq 0$ and $S_{22} \neq 0$ with S given by (3.2). Then a consistent estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ that solves the estimating equation $G_n(\theta) = 0$ exists and is unique in any compact subset of Θ containing θ_0 with a probability that goes to one as $n \to \infty$.

If, moreover,

$$\partial_\alpha \partial_y^2 g_2(0, x, x; \theta) = 0, \tag{3.9}$$

then for a martingale estimating function or if more generally $n\Delta^{2(\kappa-1)} \to 0$,

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1(\theta_0)/S_{11}^2 & 0 \\ 0 & W_2(\theta_0)/S_{22}^2 \end{pmatrix} \right) \tag{3.10}$$

where

$$\begin{aligned} W_1(\theta) &= \int_\ell^r v(x;\beta_0)[\partial_y g_1(0,x,x;\theta)]^2 \mu_{\theta_0}(x)dx = V(\theta)_{11} \\ W_2(\theta) &= \tfrac{1}{2}\int_\ell^r [v(x;\beta_0)^2 + \frac{1}{2}(v(x;\beta_0) - v(x;\beta))^2][\partial_y^2 g_2(0,x,x;\theta)]^2 \mu_{\theta_0}(x)dx. \end{aligned}$$

Note that $W_2(\theta_0) = \frac{1}{2}\int_\ell^r v(x;\beta_0)^2[\partial_y^2 g_2(0,x,x;\theta_0)]^2 \mu_{\theta_0}(x)dx$.

Thus Jacobsen's condition (1.3) and the additional condition (3.9) imply rate optimal estimators and that the estimator of the drift parameter is asymptotically independent of the estimator of the diffusion coefficient parameter. In the next section we shall see that (3.9) is automatically satisfied for efficient estimating functions. Note that for non-martingale estimating functions Δ_n must go faster to zero than was required in Theorem 3.1. Note also that if the first coordinate of g satisfies Jacobsen's condition too, then the first part of the identifiability condition in Theorem 3.3 does not hold, and the parameter α cannot be consistently estimated by the estimating function (1.2).

Like the previous theorem, Theorem 3.3 follows by asymptotic statistical results for stochastic processes, see e.g. Jacod & Sørensen (2010). To see this we need the following lemma.

Lemma 3.4 *Under the Conditions 2.1, 2.2 and 1.1*

$$D_n \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^T D_n \xrightarrow{P_{\theta_0}} \begin{pmatrix} W_1(\theta) & 0 \\ 0 & W_2(\theta) \end{pmatrix} \tag{3.11}$$

uniformly when θ is in a compact set, where

$$D_n = \begin{pmatrix} \frac{1}{\sqrt{n\Delta_n}} & 0 \\ 0 & \frac{1}{\Delta_n\sqrt{n}} \end{pmatrix}. \tag{3.12}$$

For a martingale estimating function or if more generally $n\Delta^{2(\kappa-1)} \to 0$,

$$\begin{pmatrix} \frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \\ \frac{1}{\Delta_n\sqrt{n}} \sum_{i=1}^n g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1(\theta_0) & 0 \\ 0 & W_2(\theta_0) \end{pmatrix} \right). \tag{3.13}$$

If, in addition, condition (3.9) holds, then

$$\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^{n}\partial_\alpha g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \xrightarrow{P_{\theta_0}} 0 \tag{3.14}$$

uniformly when θ is in a compact set.

Example 3.5 Consider a *quadratic martingale estimating function* of the form

$$g(\Delta, y, x;\theta) = \begin{pmatrix} a_1(x,\Delta;\theta)[y - F(\Delta,x;\theta)] \\ a_2(x,\Delta;\theta)\left[(y-F(\Delta,x;\theta))^2 - \phi(\Delta,x;\theta)\right] \end{pmatrix}, \tag{3.15}$$

where $F(\Delta,x;\theta) = E_\theta(X_\Delta|X_0 = x)$ and $\phi(\Delta,x;\theta) = \mathrm{Var}_\theta(X_\Delta|X_0 = x)$. Since, by (2.15), $F(\Delta,x;\theta) = x + O(\Delta)$ and $\phi(\Delta,x;\theta) = O(\Delta)$, we find that

$$g(0,y,x;\theta) = \begin{pmatrix} a_1(x,0;\theta)(y-x) \\ a_2(x,0;\theta)(y-x)^2 \end{pmatrix}. \tag{3.16}$$

Since $\partial_y g_2(0,y,x;\theta) = 2a_2(x,\Delta;\theta)(y-x)$, Jacobsen's condition (1.3) is satisfied. Thus estimators obtained from (3.15) are rate optimal, provided that (3.9) is satisfied, which is the case when a_2 does not depend on α.

It is illuminating to give an example of an estimating function for which estimators are not rate optimal. For

$$g(\Delta, y, x;\theta) = \begin{pmatrix} a_1(x,\Delta;\theta)[y - F(\Delta,x;\theta)] \\ a_2(x,\Delta;\theta)\left[y^2 - (\phi(\Delta,x;\theta) + F(\Delta,x;\theta)^2)\right] \end{pmatrix}, \tag{3.17}$$

we see that $\partial_y g_1(0,x,x;\theta) = a_1(x,0;\theta)$ and $\partial_y g_2(0,x,x;\theta) = a_2(x,0;\theta)2y$. The only way a linear combination of these two function can equal zero identically is if $a_1(x,0;\theta)$ is proportional to $xa_2(x,0;\theta)$. In all other cases, the estimating function given by (3.17) is not rate optimal.

□

4 Efficient estimating functions

In this section we discuss conditions under which an approximate martingale estimating function, $G_n(\theta)$, gives an efficient estimator. In particular, we show that Condition 1.2 ensures efficiency, and that estimators from Godambe-Heyde optimal estimating functions are rate optimal and efficient. The following theorem follows from Theorem 4.1 in Gobet (2002), who proved that the diffusion model (1.1) is locally asymptotically normal with Fisher information matrix equal to the inverse of $\Sigma(\theta_0)$ given by (4.1).

Theorem 4.1 *Suppose the conditions of Theorem 3.3 are satisfied. If also Condition 1.2 holds, then the estimating function (1.2) is efficient. Under (1.4) and (1.5), the asymptotic covariance matrix of the estimator $\hat\theta_n = (\hat\alpha_n, \hat\beta_n)$ is*

$$\Sigma(\theta_0) = \begin{pmatrix} \left(\int_\ell^r \frac{(\partial_\alpha b(x;\alpha_0))^2}{v(x;\beta_0)}\mu_{\theta_0}(x)dx\right)^{-1} & 0 \\ 0 & 2\left(\int_\ell^r \left[\frac{\partial_\beta v(x;\beta_0)}{v(x;\beta_0)}\right]^2 \mu_{\theta_0}(x)dx\right)^{-1} \end{pmatrix}. \tag{4.1}$$

A consistent estimator of the asymptotic variance is given by

$$\frac{1}{n\Delta_n}\sum_{i=1}^{n} g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n)^2 \xrightarrow{P_{\theta_0}} \int_\ell^r \frac{(\partial_\alpha b(x;\alpha_0))^2}{v(x;\beta_0)}\mu_{\theta_0}(x)dx$$

and

$$\frac{1}{n\Delta_n^2}\sum_{i=1}^{n} g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \hat{\theta}_n)^2 \xrightarrow{P_{\theta_0}} \int_\ell^r \left[\frac{\partial_\beta v(x;\beta_0)}{v(x;\beta_0)}\right]^2 \mu_{\theta_0}(x)dx.$$

Note that for an efficient estimating function the condition (3.9) in Theorem 3.3 is automatically satisfied, cf. (1.5).

The result is not an only-if statement because of the previously mentioned fact that different versions of the estimating function give the same estimator, but cannot all satisfy (1.4) and (1.5), even if the estimator is efficient. A martingale estimating function is efficient if and only if there exists a version that satisfies (1.4), (1.5) and the necessary previous conditions. It may typically be necessary to first multiply it by a matrix M_n depending on Δ_n. Examples of this will be given below.

The covariance matrix (4.1) is, as one would expect, equal to the leading term in the expansion of the asymptotic variance of the maximum likelihood estimator in powers of Δ found by Dacunha-Castelle & Florens-Zmirou (1986). It equals the asymptotic covariance matrix of the maximum likelihood estimator based on continuous time observation, see e.g. Kutoyants (2004). In the case of continuous time observation, the parameter β must necessarily be known. Finally, and again not a surprise, the conditions (1.4) and (1.5) are exactly the conditions for small Δ-optimality of martingale estimating functions given by Jacobsen (2001), except that he included (1.3) as part of the condition for small Δ-optimality of $\hat{\beta}$, while here it is a condition for rate optimality. Thus we have given an interpretation of the concept of small Δ-optimality in terms of the classical statistical concepts of rate optimality and efficiency. It follows from Theorem 2 in Jacobsen (2001) that (1.4) and (1.5) also imply small Δ-optimality of the approximate martingale estimating functions considered in the present paper.

Example 4.2 Consider again the *quadratic martingale estimating function* (3.15). The function $g(0, y, x; \theta)$, given by (3.16), satisfies the conditions for efficiency (1.4) and (1.5) if we choose $a_1(x, \Delta; \theta) = \partial_\alpha b(x;\alpha)/v(x;\beta)$ and $a_2(x, \Delta; \theta) = \partial_\beta v(x;\beta)/v^2(x;\beta)$, as proposed by Bibby & Sørensen (1995) and Bibby & Sørensen (1996). The same is true of any specification of the weight functions a_1 and a_2 that converge to $\partial_\alpha b/v$ and $\partial_\beta v/v^2$ as $\Delta \to 0$. An example is the optimal martingale estimating function in the sense of Godambe & Heyde (1987) (after multiplication of the second coordinate by Δ), see the papers cited above.

A similar example is obtained from the pseudo-likelihood function, where the transition density $p(\Delta, y, x; \theta)$ is replaced by the Gaussian density $\tilde{p}(\Delta, y, x; \theta)$ with mean $F(\Delta, x; \theta)$ and variance $\phi(\Delta, x; \theta)$

$$\tilde{L}_n(\theta) = \prod_{i=1}^{n} \tilde{p}(\Delta, X_{t_i^n}, X_{t_{i-1}^n}; \theta). \tag{4.2}$$

The exact conditional moments are used to ensure that a consistent estimator is obtained also in case Δ is not small. Since

$$\begin{aligned} \partial_\alpha \log \tilde{p}(\Delta, y, x; \theta) &= \frac{\partial_\alpha F(\Delta, x; \theta)}{\phi(\Delta, x; \theta)} [y - F(\Delta, x; \theta)] \\ \Delta\, \partial_\beta \log \tilde{p}(\Delta, y, x; \theta) &= \frac{\partial_\beta \phi(\Delta, x; \theta)}{\phi(\Delta, x; \theta)^2} \left[(y - F(\Delta, x; \theta))^2 - \phi(\Delta, x; \theta) \right], \end{aligned}$$

we see that the pseudo-score $\partial_\theta \log \tilde{L}_n(\theta)$ is an efficient quadratic martingale estimating function.

Clearly (3.16) holds if F and ϕ are replaced in (3.15) by expansions of order $x + O(\Delta)$ and $O(\Delta)$, respectively, so also in this non-martingale case, rate optimal estimators are obtained, provided that Δ_n goes sufficiently fast to zero. The simplest example is

$$g(\Delta, y, x; \theta) = \begin{pmatrix} a_1(x, \Delta; \theta)[y - x - b(x; \alpha)\Delta] \\ a_2(x, \Delta; \theta) \left[(y - x - b(x; \alpha)\Delta)^2 - v(\Delta, x; \beta)\Delta \right] \end{pmatrix}, \tag{4.3}$$

which gives rate optimal estimators provided that $n\Delta^2 \to 0$.

A pseudo-likelihood function can be obtained from the *Euler approximation* by replacing $\tilde{p}$ in (4.2) by

$$q(\Delta, y, x; \theta) = \frac{1}{\sqrt{2\pi v(x; \beta)\Delta}} \exp\left(-\frac{(y - x - b(x; \alpha)\Delta)^2}{2v(x; \beta)\Delta} \right).$$

The corresponding pseudo score, and hence the Euler pseudo maximum likelihood estimator, is efficient because $g(\Delta, y, x; \theta) = \partial_\theta \log q(\Delta, y, x; \theta)$, is of the form (4.3) with $a_1(x; \theta) = \partial_\alpha b(x; \alpha)/v(x; \beta)$ and (after multiplication by 2Δ) $a_2(x; \theta) = \partial_\beta v(x; \beta)/v(x; \beta)^2$. This estimator has often been used in empirical work in finance. In a similar way, it follows that the estimators considered by Prakasa Rao (1988), Florens-Zmirou (1989), Yoshida (1992), Kessler (1997) and Kelly, Platen & Sørensen (2004) are efficient under suitable conditions on the rate of convergence of Δ_n.

□

Example 4.3 Finally we consider *maximum likelihood estimation.* In broad generality, the score function is a martingale estimating function, see e.g. Barndorff-Nielsen & Sørensen (1994). The transition density can, under weak regularity conditions, be expanded in powers of Δ

$$p(\Delta, y, x; \theta) = r(\Delta, y, x; \theta)(1 + O(\Delta)),$$

where

$$r(\Delta, y, x; \theta) = \frac{1}{\sqrt{2\pi v(x; \beta)\Delta}} \exp\left(-\frac{(f(y; \beta) - f(x; \beta))^2}{2\Delta} + A(y) - A(x) - \tfrac{1}{2} \log\left(\frac{\sigma(y; \beta)}{\sigma(x; \beta)} \right) \right),$$

$f(x; \beta) = \int^x \sigma^{-1}(z; \beta) dz$ and $A(x) = \int^x b(z; \alpha)/v(z; \beta) dz$, see e.g. Dacunha-Castelle & Florens-Zmirou (1986) or Gihman & Skorohod (1972), Chapter 13. Therefore, under regularity conditions on the remainder term that need not worry us here, the score function given by $g_1(\Delta, y, x; \theta) = \partial_\alpha \log p(\Delta, y, x; \theta)$ and $g_2(\Delta, y, x; \theta) = \Delta \partial_\beta \log p(\Delta, y, x; \theta)$ satisfies that

$$\begin{aligned} g_1(0, y, x; \theta) &= \int_x^y \frac{\partial_\alpha b(z; \alpha)}{v(z; \beta)} dz + O(\Delta) \\ g_2(0, y, x; \theta) &= -[f(y; \beta) - f(x; \beta)][\partial_\beta f(y; \beta) - \partial_\beta f(x; \beta)] + O(\Delta). \end{aligned}$$

From these expansions it follows easily that the score functions (normalized as above) satisfies the Jacobsen's condition (1.3) as well as the conditions for efficiency (1.4) and (1.5). In particular, $\partial_y^2 g_2(0, x, x; \theta) = -2\partial_x f(x; \beta)\partial_\beta \partial_x^2 f(x; \beta) = \partial_\beta v(x; \beta)/v(x; \beta)^2$. Obviously, the pseudo-likelihood function obtained by replacing $\tilde{p}$ in (4.2) by r is also rate optimal and efficient provided that $n\Delta^2 \to 0$.

□

The fact that the approximate martingale estimating functions that are rate optimal and efficient are exactly those that are small Δ-optimal in the sense of Jacobsen (2001) implies that we can take advantage of the very thorough study of when martingale estimating functions satisfy the conditions (1.3), (1.4) and (1.5) presented in Jacobsen (2002). Consider martingale estimating functions of the form (2.14). It is convenient to write this type of estimating function in the following compact form

$$G_n(\theta) = \sum_{i=1}^{n} A(X_{t_{i-1}^n}, \Delta; \theta)[f(X_{t_i^n}; \theta) - \pi_\theta^\Delta f(X_{t_{i-1}^n}; \theta)], \tag{4.4}$$

where $f(y; \theta) = (f_1(y; \theta), \dots, f_N(y; \theta))^T$, $A(x, \Delta; \theta)$ a $2 \times N$-matrix of weights, and where π_θ^Δ denotes the transition operator given by

$$\pi_\theta^\Delta f(x; \theta) = E_\theta(f(X_\Delta; \theta) \mid X_0 = x) \tag{4.5}$$

The following theorem follows immediately from Theorem 2.2 of Jacobsen (2002). It is clear from the proof of this theorem that the following result holds not only for martingale estimating functions, but also for approximate martingale estimating functions satisfying (2.8).

Theorem 4.4 *Suppose Condition 2.1 is satisfied, that $N \geq 2$, and that the functions f_j are twice continuously differentiable and satisfies that the matrix*

$$D(x) = \begin{pmatrix} \partial_x f_1(x; \theta) & \partial_x^2 f_1(x; \theta) \\ \partial_x f_2(x; \theta) & \partial_x^2 f_2(x; \theta) \end{pmatrix} \tag{4.6}$$

is invertible for μ_θ-almost all x. Then a specification of the weight matrix $A(x, \Delta; \theta)$ exists such that the estimating function (4.4) satisfies the conditions (1.3), (1.4) and (1.5). When $N = 2$, these conditions are satisfy for

$$A(x, 0; \theta) = \begin{pmatrix} \partial_\alpha b(x; \alpha)/v(x; \beta) & c(x; \theta) \\ 0 & \partial_\beta v(x; \beta)/v(x; \beta)^2 \end{pmatrix} D(x)^{-1} \tag{4.7}$$

for any function $c(x; \theta)$.

Since we can index the functions f_j as we like, the condition only says that there are two functions among $f_1, \ldots, f_N$ such that D is invertible. Note also that for $N = 2$, a simple choice for the weight matrix A is to let it equal the expression in (4.7) for all Δ.

A useful way of choosing the weight matrix A in a martingale estimating function of the type (4.4) is to chose the weights that are optimal in the sense of Godambe & Heyde (1987), see also Heyde (1997). In this way we obtain estimators that minimize the asymptotic variance of estimators within the class (4.4) for a fixed, possibly large, Δ. The next theorem shows that the Godambe-Heyde optimal estimators are rate optimal and efficient in the high frequency asymptotic considered the present paper. A weight matrix A^* is Godambe-Heyde optimal if

$$\begin{aligned} A^*(x,\Delta;\theta)\, E_\theta\left([f(X_\Delta;\theta) - \pi_\theta^\Delta f(x;\theta)][f(X_\Delta;\theta) - \pi_\theta^\Delta f(x;\theta)]^T \mid X_0 = x\right) \\ = \partial_\theta \pi_\theta^\Delta f^T(x;\theta) - \pi_\theta^\Delta \partial_\theta f^T(x;\theta). \end{aligned} \tag{4.8}$$

It follows from Theorem 2.3 in Jacobsen (2002) that if $N = 2$ and the matrix D is invertible, then the Godambe-Heyde optimal martingale estimating function is rate optimal and efficient. If $N = 1$ the Godambe-Heyde optimal martingale estimating function can only be efficient if the diffusion coefficient is known, so that only the drift depends on a parameter. Here we prove that for general $N \geq 2$ the Godambe-Heyde optimal martingale estimating function is rate optimal and efficient provided that the matrix D is invertible. This result was conjectured by Jacobsen (2002) (phrased in terms of the concept small Δ-optimality).

Theorem 4.5 *Suppose Condition 2.1 is satisfied, that the functions f_j are six times continuously differentiable, that $N \geq 2$ and that the 2×2 matrix $D(x)$ given by (4.6) is invertible for μ_θ-almost all x. Let $A^*(x,\Delta;\theta)$ satisfy (4.8), and define*

$$g^*(\Delta, y, x; \theta) = \begin{pmatrix} 1 & 0 \\ 0 & 2\Delta \end{pmatrix} A^*(x,\Delta;\theta)[f(y;\theta) - \pi_\theta^\Delta f(x;\theta)]. \tag{4.9}$$

Then $g^*(0, y, x; \theta)$ satisfies (1.3), (1.4) and (1.5), and hence the estimators are rate optimal and efficient.

The fact that a condition for efficiency is $N \geq 2$ may explain the finding in Larsen & Sørensen (2007) that an optimal martingale estimating function based on two eigenfunctions seemed to be efficient for weekly observations of exchange rates in a target zone.

The efficient estimating function given by (4.9) can be used to derive simpler, equally efficient, martingale estimating functions by expanding the conditional moments in (4.8) using (2.15). Further simplification can be obtained by expanding $\pi_\theta^\Delta f(x;\theta)$ in (4.9).

Let us conclude this section by stating the results for a d-dimensional diffusion. In this case $b(x;\alpha)$ is d-dimensional and $v(x;\beta) = \sigma(x;\beta)\sigma(x;\beta)^T$ is a $d \times d$-matrix. The conditions for efficiency are

$$\partial_y g_1(0, x, x; \theta) = \partial_\alpha b(x;\alpha)^T v(x;\beta)^{-1}$$

and

$$\text{vec}\left(\partial_y^2 g_2(0, x, x; \theta)\right) = \text{vec}\left(\partial_\beta v(x;\beta)\right)\left(v^{\otimes 2}(x;\beta)\right)^{-1}.$$

In the latter equation, $\text{vec}(M)$ denotes for a $d \times d$ matrix M the d^2-dimensional row vector consisting of the rows of M placed one after the other, and $M^{\otimes 2}$ is the $d^2 \times d^2$-matrix with

$(i',j'),(ij)$th entry equal to $M_{i'i}M_{j'j}$. Thus if $M = \partial_\beta v(x;\beta)$ and $M^\bullet = (v^{\otimes 2}(x;\beta))^{-1}$, then the (i,j)th coordinate of $\mathrm{vec}(M)\, M^\bullet$ is $\sum_{i'j'} M_{i'j'} M^\bullet_{(i'j'),(i,j)}$. These expressions are the conditions for small Δ-optimality for multivariate diffusions given by Jacobsen (2002).

For a d-dimensional diffusion process, the condition analogous to the one in Theorem 4.4 ensuring the existence of a rate optimal and efficient estimating function of the form (4.4) is that $N \geq d(d+3)/2$, and that the $N \times (d+d^2)$-matrix

$$\begin{pmatrix} \partial_x f(x;\theta) & \partial_x^2 f(x;\theta) \end{pmatrix}$$

has full rank $d(d+3)/2$, see Jacobsen (2002). When α and β are multivariate, we further need that $\{\partial_{\alpha_i} b(x;\alpha)\}$ and $\{\partial_{\beta_i} v(x;\beta)\}$ are two sets of linearly independent functions of x. These conditions also ensure that Theorem 4.5 holds for a d-dimensional diffusion process, i.e. that the Godambe-Heyde optimal martingale estimating function is rate optimal and efficient for a d-dimensional diffusion process.

5 Proofs

The first two lemmas are essentially Lemma 6 and Lemma 8 in Kessler (1997). The reader is reminded that $R(\Delta, y, x;\theta)$ denotes a (generic) function such that $|R(\Delta, y, x;\theta)| \leq F(y, x;\theta)$ where F is of polynomial growth in y and x uniformly for θ in compact sets. Similarly for $R(\Delta, x;\theta)$. We sometimes use the notation $a \leq_C b$, which means that there exists a $C > 0$ such that $a \leq Cb$.

Lemma 5.1 *Assume Condition 2.1. For $k = 1, 2, \ldots$ a constant $C_k > 0$ exists such that*

$$E_{\theta_0}(|X_{t+\Delta} - X_t|^k \mid X_t) \leq C_k \Delta^{k/2}(1 + |X_t|)^{C_k} \tag{5.1}$$

for $\Delta > 0$. Let $f(y, x, \theta)$ be a real function of polynomial growth in x and y uniformly for θ in a compact set K. Then there exists a constant $C > 0$ such that for any fixed $\Delta_0 > 0$

$$E_{\theta_0}(|f(X_{t+\Delta}, X_t, \theta)| \mid X_t) \leq C(1 + |X_t|)^C \quad \text{for } \Delta \in [0, \Delta_0] \text{ and } \theta \in K. \tag{5.2}$$

Suppose the function $f(y, x, \theta)$ is, moreover, $2k$ times differentiable ($k = 0, 1, 2, 3$) with respect to y with derivatives of polynomial growth in x and y uniformly for θ in compact sets. Then

$$\int_0^\Delta \int_0^{u_1} \cdots \int_0^{u_{k-1}} E_{\theta_0}\left(L_{\theta_0}^k f(X_{t+u_k}, X_t, \theta) \mid X_t\right) du_k \cdots du_1 = \Delta^k R(\Delta, X_t, \theta). \tag{5.3}$$

The result (5.3) is used to ensure that the remainder term in expansions of the type (2.15) have the expected order. It could be proved for larger values of k if stronger conditions were imposed on the coefficients b and σ.

Lemma 5.2 *Assume Condition 2.1, and let $f(x, \theta)$ be a real function that is differentiable with respect to x and θ with derivatives of polynomial growth in x uniformly for θ in a compact set. Then*

$$\frac{1}{n}\sum_{i=1}^n f(X_{t_i^n}, \theta) \xrightarrow{P_{\theta_0}} \int_\ell^r f(x, \theta)\mu_{\theta_0}(x)dx$$

uniformly for θ in a compact set.

Lemma 9 in Genon-Catalot & Jacod (1993) is used frequently in the proofs of Lemma 3.2 and Lemma 3.4 to establish pointwise convergence. The result is therefore cited here for the convenience of the reader.

Lemma 5.3 *Let Z_i^n $(i = 1, \ldots, n, n \in \mathbb{N})$ be a triangular array of random variables such that Z_i^n is $\mathcal{G}_i^n$-measurable, where $\mathcal{G}_i^n = \sigma(W_s : s \leq t_i^n)$. If*

$$\sum_{i=1}^{n} E_\theta(Z_i^n \,|\, \mathcal{G}_{i-1}^n) \xrightarrow{P_\theta} U$$

and

$$\sum_{i=1}^{n} E_\theta((Z_i^n)^2 \,|\, \mathcal{G}_{i-1}^n) \xrightarrow{P_\theta} 0,$$

where U is a random variable, then

$$\sum_{i=1}^{n} Z_i^n \xrightarrow{P_\theta} U.$$

Proof of Lemma 2.3. Combining (2.8), (2.9) and (2.15) and using Lemma 5.1, we find that

$$\begin{aligned}
O(\Delta^\kappa) &= E_\theta(g(\Delta, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \,|\, X_{t_{i-1}^n}) \\
&= g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) + \Delta \left[L_\theta(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \right] \\
&+ \tfrac{1}{2}\Delta^2 \left[L_\theta^2(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + 2L_\theta(g^{(1)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + g^{(2)}(X_{t_{i-1}^n}, X_{t_{i-1}^n}; \theta) \right] \\
&+ \Delta^3 R(\Delta, X_{t_{i-1}^n}, \theta),
\end{aligned}$$

from which the lemma follows.

□

In order to establish uniform convergence in the proofs of Lemma 3.2 and Lemma 3.4, we need a technical lemma, which is easier to formulate with the following condition.

Condition 5.4 *The real function $f(\Delta, y, x; \theta)$ satisfies that $f(0, x, x; \theta) = 0$ for all $x \in (\ell, r)$ and $\theta \in \Theta$, and $f \in C_{p,1,2,1}(\mathbb{R}_+, (\ell, r)^2, \Theta)$.*

Lemma 5.5 *Assume Condition 2.1, and let $f(\Delta, y, x; \theta)$ be a function that satisfies Condition 5.4. Then for every $m \in \mathbb{N}$ and for every compact $K \subseteq \Theta$, a constant $C_{m,K} > 0$ exists such that*

$$E_{\theta_0}\left(|\zeta_n(\theta_2) - \zeta_n(\theta_1)|^{2m}\right) \leq C_{m,K}|\theta_2 - \theta_1|^{2m} \tag{5.4}$$

for all θ_1 and θ_2 in K and for all n, where

$$\zeta_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^{n} f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta). \tag{5.5}$$

If, moreover, the functions h_1 and h_2 given by

$$\begin{aligned} h_1(s,y,x;\theta) &= \partial_s f(s,y,x;\theta) + \partial_y f(s,y,x;\theta)b(y;\alpha_0) + \tfrac{1}{2}\partial_y^2 f(s,y,x;\theta)v(y,\beta_0) \\ h_2(s,y,x;\theta) &= \partial_y f(s,y,x;\theta)\sigma(y;\beta_0). \end{aligned}$$

satisfy Condition 5.4, then a constant $C_{m,K} > 0$ exists such that

$$E_{\theta_0}\left(|\phi_n(\theta_2) - \phi_n(\theta_1)|^{2m}\right) \leq C_{m,K}|\theta_2 - \theta_1|^{2m} \tag{5.6}$$

for all θ_1 and θ_2 in the compact set K and for all n, where

$$\phi_n(\theta) = \frac{1}{n\Delta_n^{3/2}} \sum_{i=1}^n f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta). \tag{5.7}$$

Finally, if the functions

$$h_{i2}(s,y,x;\theta) = \partial_y h_i(s,y,x;\theta)\sigma(y;\beta_0), \quad i = 1,2, \tag{5.8}$$

satisfy Condition 5.4, then a constant $C_{m,K} > 0$ exists such that

$$E_{\theta_0}\left(|\xi_n(\theta_2) - \xi_n(\theta_1)|^{2m}\right) \leq C_{m,K}|\theta_2 - \theta_1|^{2m} \tag{5.9}$$

for all θ_1 and θ_2 in the compact set K and for all n, where

$$\xi_n(\theta) = \frac{1}{n\Delta_n^2} \sum_{i=1}^n f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta). \tag{5.10}$$

Proof. By Ito's formula

$$f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) = \int_{t_{i-1}^n}^{t_i^n} h_1(s, X_s, X_{t_{i-1}^n}; \theta) ds + \int_{t_{i-1}^n}^{t_i^n} h_2(s, X_s, X_{t_{i-1}^n}; \theta) dW_s, \tag{5.11}$$

By Condition 5.4, the partial derivatives $\partial_\theta h_1$ and $\partial_\theta h_2$ are of polynomial growth in y and x uniformly for θ in a compact set. We can treat the two terms on the right hand side of (5.11) separately. Define $Dh_i(s,y,x;\theta_2,\theta_1) = h_i(s,y,x;\theta_2) - h_i(s,y,x;\theta_1)$. Using Jensen's inequality twice, we obtain

$$\begin{aligned} &\frac{1}{\Delta_n^{2m}} E_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} Dh_1(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1) ds\right|^{2m}\right) \\ &\leq \quad \frac{1}{n\Delta_n^{2m}} \sum_{i=1}^n E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{t_i^n} Dh_1(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1) ds\right|^{2m}\right) \\ &\leq \quad \frac{1}{n\Delta_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(|Dh_1(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1)|^{2m}\right) ds \\ &\leq_C \quad \frac{1}{n\Delta_n} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(\left|\int_0^1 \partial_\theta h_1(s, X_s, X_{t_{i-1}^n}; \theta_1 + u(\theta_2 - \theta_1)) du\right|^{2m}\right) ds |\theta_2 - \theta_1|^{2m} \\ &\leq_C \quad |\theta_2 - \theta_1|^{2m}. \end{aligned}$$

By the Burkholder-Davis-Gundy inequality and Jensen's inequality

$$\begin{aligned}
&\frac{1}{\Delta_n^{2m}} E_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} Dh_2(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1) dW_s\right|^{2m}\right) \\
\leq_C \quad & \frac{1}{\Delta_n^{2m}} E_{\theta_0}\left(\left|\frac{1}{n^2}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} Dh_2(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1)^2 ds\right|^{m}\right) \\
\leq \quad & \frac{1}{n^{m+1}\Delta_n^{2m}} \sum_{i=1}^n E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{t_i^n} Dh_2(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1)^2 ds\right|^{m}\right) \\
\leq \quad & \frac{1}{(n\Delta_n)^{m+1}} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(|Dh_2(s, X_s, X_{t_{i-1}^n}; \theta_2, \theta_1)|^{2m}\right) ds \\
\leq_C \quad & \frac{1}{(n\Delta_n)^m} |\theta_2 - \theta_1|^{2m}.
\end{aligned}$$

The results (5.6) and (5.9) follow in a similar way. Under the conditions for (5.6),

$$\begin{aligned}
&f(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) = \\
&\int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} h_{11}(u, X_u, X_{t_{i-1}^n}; \theta) du ds + \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} h_{12}(u, X_u, X_{t_{i-1}^n}; \theta) dW_u ds \\
&+ \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} h_{21}(u, X_u, X_{t_{i-1}^n}; \theta) du dW_s + \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} h_{22}(u, X_u, X_{t_{i-1}^n}; \theta) dW_u dW_s
\end{aligned} \tag{5.12}$$

with h_{i2} given by (5.8) and

$$h_{i1}(s, y, x; \theta) = \partial_s h_i(s, y, x; \theta) + \partial_y h_i(s, y, x; \theta) b(y; \alpha_0) + \tfrac{1}{2}\partial_y^2 h_i(s, y, x; \theta) v(y, \beta_0).$$

With Dh_{ij} defined as previously, we see that

$$\begin{aligned}
&\frac{1}{\Delta_n^{4m}} E_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} Dh_{11}(u, X_u, X_{t_{i-1}^n}; \theta_2, \theta_1) du ds\right|^{2m}\right) \\
\leq \quad & \frac{1}{n\Delta_n^{4m}} \sum_{i=1}^n E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} Dh_{11}(u, X_u, X_{t_{i-1}^n}; \theta_2, \theta_1) du ds\right|^{2m}\right) \\
\leq \quad & \frac{1}{n\Delta_n^{2}} \sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} \int_{t_{i-1}^n}^{s} E_{\theta_0}\left(|Dh_{11}(u, X_u, X_{t_{i-1}^n}; \theta_2, \theta_1)|^{2m}\right) du ds \\
\leq_C \quad & |\theta_2 - \theta_1|^{2m},
\end{aligned}$$

$$
\begin{aligned}
&\frac{1}{\Delta_n^{3m}} E_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} Dh_{12}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u ds\right|^{2m}\right)\\
&\le \frac{1}{\Delta_n^{3m} n}\sum_{i=1}^n E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} Dh_{12}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u ds\right|^{2m}\right)\\
&\le \frac{1}{\Delta_n^{m+1} n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{s} Dh_{12}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u\right|^{2m}\right) ds\\
&\le_C \frac{1}{\Delta_n^{m+1} n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{s} Dh_{12}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)^2 ds\right|^{m}\right) ds\\
&\le \frac{1}{\Delta_n^{2} n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} E_{\theta_0}\left(|Dh_{12}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)|^{2m} du ds\right)\\
&\le_C |\theta_2-\theta_1|^{2m},
\end{aligned}
$$

$$
\begin{aligned}
&\frac{1}{\Delta_n^{4m}} E_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} Dh_{21}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)du dW_s\right|^{2m}\right)\\
&\le_C \frac{1}{\Delta_n^{4m}} E_{\theta_0}\left(\left|\frac{1}{n^2}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\left(\int_{t_{i-1}^n}^{s} Dh_{21}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)du\right)^2 ds\right|^{m}\right)\\
&\le \frac{1}{n^{m+1}\Delta_n^{4m}}\sum_{i=1}^n E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{t_i^n}\left(\int_{t_{i-1}^n}^{s} Dh_{21}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)du\right)^2 ds\right|^{m}\right)\\
&\le \frac{1}{n^{m+1}\Delta_n^{m+2}}\sum_{i=1}^n \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} E_{\theta_0}\left(|Dh_{21}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)|^{2m} du ds\right)\\
&\le_C \frac{1}{(n\Delta_n)^m}|\theta_2-\theta_1|^{2m},
\end{aligned}
$$

and that

$$
\begin{aligned}
&\frac{1}{\Delta_n^{3m}} E_{\theta_0}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} Dh_{22}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u dW_s\right|^{2m}\right)\\
&\leq_C \quad \frac{1}{\Delta_n^{3m}} E_{\theta_0}\left(\left|\frac{1}{n^2}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\left(\int_{t_{i-1}^n}^{s} Dh_{22}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u\right)^2 ds\right|^{m}\right)\\
&\leq \quad \frac{1}{n^{m+1}\Delta_n^{3m}}\sum_{i=1}^{n} E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{t_i^n}\left(\int_{t_{i-1}^n}^{s} Dh_{22}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u\right)^2 ds\right|^{m}\right)\\
&\leq \quad \frac{1}{n^{m+1}\Delta_n^{2m+1}}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{s} Dh_{22}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)dW_u\right|^{2m}\right) ds\\
&\leq_C \quad \frac{1}{n^{m+1}\Delta_n^{2m+1}}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n} E_{\theta_0}\left(\left|\int_{t_{i-1}^n}^{s} Dh_{22}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)^2 du\right|^{m}\right) ds\\
&\leq \quad \frac{1}{n^{m+1}\Delta_n^{m+2}}\sum_{i=1}^{n}\int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} E_{\theta_0}\left(|Dh_{22}(u,X_u,X_{t_{i-1}^n};\theta_2,\theta_1)|^{2m} du ds\right)\\
&\leq_C \quad \frac{1}{(n\Delta_n)^m}|\theta_2-\theta_1|^{2m}.
\end{aligned}
$$

We have already taken care of two of the terms in (5.12) on the way to prove (5.9). The terms involving h_{12} and h_{22} require more work. Since h_{i2}, $i = 1, 2$ satisfy Condition 5.4, we find that

$$
\begin{aligned}
&\int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} h_{12}(u,X_u,X_{t_{i-1}^n};\theta)dW_u ds =\\
&\quad \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s}\int_{t_{i-1}^n}^{u} h_{121}(v,X_v,X_{t_{i-1}^n};\theta)dv dW_u ds + \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s}\int_{t_{i-1}^n}^{u} h_{122}(v,X_v,X_{t_{i-1}^n};\theta)dW_v dW_u ds
\end{aligned}
$$

and

$$
\begin{aligned}
&\int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s} h_{22}(u,X_u,X_{t_{i-1}^n};\theta)dW_u dW_s =\\
&\quad \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s}\int_{t_{i-1}^n}^{u} h_{221}(v,X_v,X_{t_{i-1}^n};\theta)dv dW_u dW_s + \int_{t_{i-1}^n}^{t_i^n}\int_{t_{i-1}^n}^{s}\int_{t_{i-1}^n}^{u} h_{222}(v,X_v,X_{t_{i-1}^n};\theta)dW_v dW_u dW_s,
\end{aligned}
$$

where

$$
\begin{aligned}
h_{i21}(s,y,x;\theta) &= \partial_\Delta h_{i2}(s,y,x;\theta) + \partial_y h_{i2}(s,y,x;\theta)b(y;\alpha_0) + \tfrac{1}{2}\partial_y^2 h_{i2}(s,y,x;\theta)v(y,\beta_0)\\
h_{i22}(s,y,x;\theta) &= \partial_y h_{i2}(s,y,x;\theta)\sigma(y;\beta_0).
\end{aligned}
$$

The result is now obtained by evaluating the triple integrals using the Burkholder-Davis-Gundy inequality and Jensen's inequality exactly as above. □

Proof of Lemma 3.2. By (2.9), (2.15), (2.11) and Lemma 5.1,

$$
\begin{aligned}
& E_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&= \Delta_n \left[g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + L_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right] + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}, \theta) \\
&= \Delta_n \left[L_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - L_{\theta}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right] + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}, \theta).
\end{aligned}
$$

The last equality follows from (2.12). Thus

$$
\begin{aligned}
& \frac{1}{n\Delta_n}\sum_{i=1}^n E_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \\
&= \frac{1}{n}\sum_{i=1}^n \left[L_{\theta_0}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - L_{\theta}(g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right] + \Delta_n \frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}, \theta) \\
& \xrightarrow{P_{\theta_0}} \gamma(\theta, \theta_0)
\end{aligned}
$$

by Lemma 5.2. Moreover, $E_{\theta_0}\left(g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right) = \Delta_n R(\Delta_n, X_{t_{i-1}^n}, \theta)$, so

$$
\frac{1}{(n\Delta_n)^2}\sum_{i=1}^n E_{\theta_0}\left(g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2 \mid X_{t_{i-1}^n}\right) = \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}, \theta) \xrightarrow{P_{\theta_0}} 0.
$$

Therefore pointwise convergence in (3.5) follows from Lemma 5.3. In order to prove that the convergence is uniform for θ in a compact set K, we show that the sequence $\zeta_n(\cdot) = \frac{1}{n\Delta_n}\sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}, \cdot)$ converges weakly to the limit $\gamma(\cdot, \theta_0)$ in the space, $C(K)$, of continuous functions on K with the supremum norm. Since the limit is non-random, this implies uniform convergence in probability for $\theta \in K$. We have proved pointwise convergence, so the weak convergence result follows because the family of distributions of $\zeta_n(\cdot)$ is tight. The tightness follows from Lemma 5.5 with $f = g_i$ and $m = 2$. That (5.4) and pointwise convergence implies tightness follows from Corollary 14.9 in Kallenberg (1997), which is a generalization of Theorem 12.3 in Billingsley (1968) (see also Lemma 3.1 in Yoshida (1990) and Theorem 20 in Appendix I of Ibragimov & Has'minskii (1981)).

In a similar way it follows from (2.9), (2.15), (2.11), (2.12) and Lemma 5.1 that

$$
\begin{aligned}
& E_{\theta_0}\left(\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) \mid X_{t_{i-1}^n}\right) \qquad (5.13) \\
&= \Delta_n \left[\partial_\theta g^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta) + L_{\theta_0}(\partial_\theta g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right] + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}, \theta) \\
&= \Delta_n \left[L_{\theta_0}(\partial_\theta g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - L_{\theta}(\partial_\theta g(0;\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n}) - A_\theta(X_{t_{i-1}^n})\right] \\
&\qquad + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}, \theta),
\end{aligned}
$$

and from (2.9),(2.15), (2.11), and Lemma 5.1 that

$$
\begin{aligned}
& E_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta) g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^T \mid X_{t_{i-1}^n}\right) \\
&= \Delta_n v(X_{t_{i-1}^n}, \beta_0)\partial_y g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)\partial_y g(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^T + \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n}, \theta).
\end{aligned}
$$

Since by (2.9),(2.15), (2.11), and Lemma 5.1

$$E_{\theta_0}\left([\partial_\theta g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta)]^2 \mid X_{t_{i-1}^n}\right) = \Delta_n R(\Delta_n, X_{t_{i-1}^n}, \theta)$$

and

$$E_{\theta_0}\left([g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) g_k(\Delta, X_{t_i^n}, X_{t_{i-1}^n}; \theta)]^2 \mid X_{t_{i-1}^n}\right) = \Delta_n R(\Delta_n, X_{t_{i-1}^n}, \theta), \tag{5.14}$$

we can, as above, use Lemma 5.2 and Lemma 5.3 to prove (3.6) and (3.7). As above, uniform convergence for θ in a compact set K follows by using Lemma 5.5 with $f = \partial_{\theta_j} g_k$ and $f = g_j g_k$ to prove the tightness of (5.5) in $C(K)$.

Finally, (3.8) follows from the central limit theorem for square integrable martingale arrays under conditions which, in the martingale case, we have already verified in the proof of (3.7), see e.g. Corollary 3.1 in Hall & Heyde (1980) with the conditional Lindeberg condition replaced by the stronger conditional Liapounov condition that follows from (5.14) and Lemma 5.2, e.g.

$$\frac{1}{(n\Delta_n)^2} \sum_{i=1}^n E_{\theta_0}\left(g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)^4 \mid X_{t_{i-1}^n}\right) = \frac{1}{n\Delta_n} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}, \theta_0) \xrightarrow{P_{\theta_0}} 0.$$

The nestedness condition in Hall and Heyde's Corollary 3.1 is not needed here because the limit of the quadratic variation is non-random. In the case of non-martingale estimating functions, we also need that by (2.8)

$$\frac{1}{\sqrt{n\Delta_n}} \sum_{i=1}^n E_{\theta_0}\left(g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0) \mid X_{t_{i-1}^n}\right) = \sqrt{n}\Delta_n^{\kappa-1/2} \frac{1}{n} \sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n}, \theta_0) \xrightarrow{P_{\theta_0}} 0, \tag{5.15}$$

and it must be checked that the martingale $\sum_{i=1}^n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)$, where $\tilde{g} = g - E_{\theta_0}\left(g \mid X_{t_{i-1}^n}\right)$, satisfies the conditions of the central limit theorem. This follows from the expansions of conditional expectations given above and $E_{\theta_0}\left(g_j(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta_0)^3 \mid X_{t_{i-1}^n}\right) = \Delta_n R(\Delta_n, X_{t_{i-1}^n}, \theta_0)$.

□

Proof of Theorem 3.1. By Lemma 3.2, the estimating function

$$G_n(\theta) = \frac{1}{n\Delta_n} \sum_{i=1}^n g(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}; \theta) \tag{5.16}$$

satisfies the conditions that $G_n(\theta_0) \xrightarrow{P_{\theta_0}} 0$, $\partial_\theta G_n(\theta) \xrightarrow{P_{\theta_0}} U(\theta)$ uniformly for θ in a compact set, and that $U(\theta_0) = -S$ is invertible, where $U(\theta)$ denotes the right hand side of (3.6). This implies the eventual existence and the consistency of $\hat{\theta}_n$ as well as the eventual uniqueness of a consistent estimator on any compact subset of Θ containing θ_0; see Jacod & Sørensen (2010). The facts that the limit of $G_n(\theta)$ satisfies that $\gamma(\theta, \theta_0) \neq 0$ for $\theta \neq \theta_0$ and is continuous in θ imply that any non-consistent solution to the estimating equation will eventually leave any compact subset of Θ containing θ_0. The asymptotic normality follows by standard arguments, see e.g. Jacod & Sørensen (2010). □

Proof of Lemma 3.4. By (2.9), (2.15), (2.11), (1.3) and Lemma 5.1,

$$\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^n E_{\theta_0}\left(g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\mid X_{t_{i-1}^n}\right) \\ = \Delta_n^{1/2}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n},\theta) \xrightarrow{P_{\theta_0}} 0$$

and

$$\frac{1}{n^2\Delta_n^3}\sum_{i=1}^n E_{\theta_0}\left([g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)]^2\mid X_{t_{i-1}^n}\right) \\ = \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n},\theta) \xrightarrow{P_{\theta_0}} 0, \tag{5.17}$$

so the pointwise convergence of the two off-diagonal entries in (3.11) follows from Lemma 5.3. Similarly to the proof of Lemma 3.2, uniform convergence for θ in a compact set K follows by using Lemma 5.5 with $f = g_1g_2$ to prove the tightness of (5.7) in $C(K)$.

The convergence of $(n\Delta_n)^{-1}\sum_{i=1}^n g_1(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2$ was taken care of in Lemma 3.2. By (2.9), (2.15), (2.11), (2.12), (1.3) and Lemma 5.1, we see that

$$\begin{aligned} & E_{\theta_0}\left(g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2\mid X_{t_{i-1}^n}\right) \\ = \;& \Delta_n^2\left[\tfrac{1}{2}L_{\theta_0}^2(g_2(0;\theta)^2)(X_{t_{i-1}^n}, X_{t_{i-1}^n}) + 2L_{\theta_0}(g_2(0;\theta)g_2^{(1)}(\theta))(X_{t_{i-1}^n}, X_{t_{i-1}^n})\right. \\ & \qquad \left. + g_2^{(1)}(X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta)^2\right] + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n},\theta) \\ = \;& \tfrac{1}{2}\Delta_n^2\left[v(X_{t_{i-1}^n};\beta_0)^2 + \tfrac{1}{2}(v(X_{t_{i-1}^n};\beta_0) - v(X_{t_{i-1}^n};\beta))^2\right](\partial_y^2 g_2(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta))^2 \\ & \qquad + \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n},\theta), \end{aligned}$$

Thus

$$\begin{aligned} & \frac{1}{n\Delta_n^2}\sum_{i=1}^n E_{\theta_0}\left(g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^2\mid X_{t_{i-1}^n}\right) \\ = \;& \frac{1}{n}\sum_{i=1}^n \tfrac{1}{2}\left[v(X_{t_{i-1}^n};\beta_0) + \tfrac{1}{2}(v(X_{t_{i-1}^n};\beta_0) - v(X_{t_{i-1}^n};\beta))^2\right](\partial_y^2 g_2(0, X_{t_{i-1}^n}, X_{t_{i-1}^n};\theta))^2 \\ & \qquad + \Delta_n\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n},\theta) \\ & \xrightarrow{P_{\theta_0}} W_2(\theta) \end{aligned}$$

by Lemma 5.2. We conclude that $(n\Delta_n^2)^{-1}\sum_{i=1}^n g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n}\theta)^2$ converges to $W_2(\theta)$ by Lemma 5.3 because

$$\frac{1}{n^2\Delta_n^4}\sum_{i=1}^n E_{\theta_0}\left(g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)^4\mid X_{t_{i-1}^n}\right) = \frac{1}{n\Delta_n}\frac{1}{n}\sum_{i=1}^n R(\Delta_n, X_{t_{i-1}^n},\theta) \xrightarrow{P_{\theta_0}} 0. \tag{5.18}$$

This follows from (2.9), (2.15), (2.11), (1.3), and Lemmas 5.1 and 5.2. Uniform convergence for θ in a compact set K follows by using Lemma 5.5 with $f = g_2^2$ to prove the tightness of (5.10) in $C(K)$.

As in the proof of Lemma 3.2, (3.13) follows from the central limit theorem for square integrable martingale arrays (Corollary 3.1 in Hall & Heyde (1980)) under conditions which, in the martingale case, we have already verified in the proof of (3.11). In particular, the conditional Liapounov condition follows from (5.14), (5.18) and (5.17). In the case of non-martingale estimating functions, we also need that g_1 satisfies (5.15) and that by (2.8)

$$\frac{1}{\Delta_n\sqrt{n}}\sum_{i=1}^{n} E_{\theta_0}\left(g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)\,|\,X_{t_{i-1}^n}\right) = \sqrt{n}\Delta_n^{\kappa-1}\frac{1}{n}\sum_{i=1}^{n} R(\Delta_n, X_{t_{i-1}^n},\theta_0) \xrightarrow{P_{\theta_0}} 0,$$

and it must be checked that the martingale $\sum_{i=1}^n \tilde{g}(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)$, where $\tilde{g} = g - E_{\theta_0}\left(g|\,X_{t_{i-1}^n}\right)$, satisfies the conditions of the central limit theorem. This follows from the expansions of conditional expectations given above and $E_{\theta_0}\left(g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta_0)^3\,|\,X_{t_{i-1}^n}\right) = \Delta_n^2 R(\Delta_n, X_{t_{i-1}^n},\theta_0)$.

Finally, to prove (3.14) note that (5.13), (1.3) and (3.9) imply that

$$E_{\theta_0}\left(\partial_\alpha g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,|\,X_{t_{i-1}^n}\right) = \Delta^2 R(\Delta_n, X_{t_{i-1}^n},\theta),$$

and that it follows from (2.9), (2.15), (2.11), (2.12), (1.3), (3.9) and Lemma 5.1 that

$$E_{\theta_0}\left([\partial_\alpha g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)]^2\,|\,X_{t_{i-1}^n}\right) = \Delta_n^3 R(\Delta_n, X_{t_{i-1}^n},\theta).$$

Therefore by Lemma 5.2

$$\frac{1}{n\Delta_n^{3/2}}\sum_{i=1}^{n} E_{\theta_0}\left(\partial_\alpha g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)\,|\,X_{t_{i-1}^n}\right) = \sqrt{\Delta_n}\frac{1}{n}\sum_{i=1}^{n} R(\Delta_n, X_{t_{i-1}^n};\theta) \xrightarrow{P_{\theta_0}} 0.$$

and

$$\frac{1}{n^2\Delta_n^3}\sum_{i=1}^{n} E_{\theta_0}\left([\partial_\alpha g_2(\Delta_n, X_{t_i^n}, X_{t_{i-1}^n};\theta)]^2\,|\,X_{t_{i-1}^n}\right) = \frac{1}{n}\frac{1}{n}\sum_{i=1}^{n} R(\Delta_n, X_{t_{i-1}^n},\theta) \xrightarrow{P_{\theta_0}} 0,$$

so that (3.14) follows from Lemma 5.3. Uniform convergence for θ in a compact set K follows by using Lemma 5.5 with $f = \partial_\alpha g_2$ to conclude tightness of (5.7) in $C(K)$. To see that $\partial_\alpha g_2$ satisfies the conditions of the lemma, we use (2.11) to conclude that $\partial_\Delta\partial_\alpha g_2(0,x,x;\theta) = \partial_\alpha g_2^{(1)}(x,x;\theta) = -\partial_\alpha L_\theta(g_2(0;\theta))(x,x) = 0$.

□

Proof of Theorem 3.3. The eventual existence and uniqueness and the consistence of $\hat{\theta}_n$ on any compact subset of Θ containing θ_0 follows from Theorem 3.1: Since (1.3) implies $S_{21} = 0$, the assumptions that $S_{11} \neq 0$ and $S_{22} \neq 0$ ensure that S is invertible, and under Condition 1.1 the identifiability condition imposed in Theorem 3.3 ensures that $\gamma(\theta,\theta_0) \neq 0$ for $\theta \neq \theta_0$ with γ, the limit of $G_n(\theta)$, given by (3.1).

To prove the asymptotic normality (3.10) of the estimator $\hat{\theta}_n$ we consider

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

where D_n is given by (3.12). On the set $\{\tilde{G}_n(\hat{\theta}_n) = 0\}$ (the probability of which goes to one)

$$-\partial_{\theta^T} \tilde{G}_n(\theta_n^{(1)}, \theta_n^{(2)}) A_n^{-1} A_n(\hat{\theta}_n - \theta_0) = \tilde{G}_n(\theta_0),$$

where

$$A_n = \begin{pmatrix} \sqrt{\Delta_n n} & 0 \\ 0 & \sqrt{n} \end{pmatrix},$$

$\partial_{\theta^T} \tilde{G}_n(\theta_n^{(1)}, \theta_n^{(2)})$ is the 2×2-matrix whose jkth entry is $\partial_{\theta_k} \tilde{G}_n(\theta_n^{(j)})_j$, and $\theta_n^{(j)}$ is a random convex combination of $\hat{\theta}_n$ and θ_0. Since by (3.6) and (3.14)

$$-\partial_{\theta^T} \tilde{G}_n(\theta_n^{(1)}, \theta_n^{(2)}) A_n^{-1} \xrightarrow{P_{\theta_0}} \begin{pmatrix} S_{11} & 0 \\ 0 & S_{22} \end{pmatrix},$$

(3.10) follows from (3.13).

□

Proof of Theorem 4.4. This theorem follows from Theorem 2.2 of Jacobsen (2002). It is, however, instructive to give a proof that when $N = 2$ and A is given by (4.7), then the estimating function (4.4) satisfies (1.3), (1.4) and (1.5). For $g(\Delta, y, x; \theta) = A(x, \Delta; \theta)[f(y; \theta) - \pi_\theta^\Delta f(x; \theta)]$,

$$\begin{aligned} \partial_y g(0, y, x; \theta) &= A(x, 0; \theta)\, \partial_y f(y) \\ \partial_y^2 g(0, y, x; \theta) &= A(x, 0; \theta)\, \partial_y^2 f(y). \end{aligned}$$

Therefore

$$\begin{aligned} \left(\partial_y g(0, x, x; \theta), \partial_y^2 g(0, x, x; \theta)\right) &= \begin{pmatrix} \partial_\alpha b(x; \alpha)/v(x; \beta) & c(x; \theta) \\ 0 & \partial_\beta v(x; \beta)/v(x; \beta)^2 \end{pmatrix} D(x)^{-1} D(x) \\ &= \begin{pmatrix} \partial_\alpha b(x; \alpha)/v(x; \beta) & c(x; \theta) \\ 0 & \partial_\beta v(x; \beta)/v(x; \beta)^2 \end{pmatrix}, \end{aligned}$$

from which we read (1.3), (1.4) and (1.5).

□

Proof of Theorem 4.5. By (2.15)

$$\pi_\theta^\Delta f(x; \theta) = f(x; \theta) + \Delta L_\theta f(x; \theta) + \tfrac{1}{2} \Delta^2 L_\theta^2 f(x; \theta) + O(\Delta^3), \tag{5.19}$$

which after another application of (2.15) implies that for $h(\Delta, y, x; \theta) = f(y; \theta) - \pi_\theta^\Delta f(x; \theta)$

$$\begin{aligned} E_\theta\left(h(\Delta, X_\Delta, x; \theta) h(\Delta, X_\Delta, x; \theta)^T \mid X_0 = x\right) &= \Delta L_\theta(h(0; \theta) h(0; \theta)^T)(x, x) \qquad (5.20) \\ &\quad + \Delta^2 \left(\tfrac{1}{2} L_\theta^2(h(0; \theta) h(0; \theta)^T)(x, x) - L_\theta f(x; \theta) L_\theta f^T(x; \theta)\right) + O(\Delta^3) \\ = \ & \Delta v(x; \beta) \partial_x f(x; \theta) \partial_x f(x; \theta)^T \\ & + \Delta^2 \left[q_1(x; \theta) \partial_x f(x; \theta) \partial_x f(x; \theta)^T + q_2(x; \theta) \left(\partial_x^2 f(x; \theta) \partial_x f(x; \theta)^T + \partial_x f(x; \theta) \partial_x^2 f(x; \theta)^T\right)\right. \\ & \left. + v(x; \beta)^2 \left(\partial_x^2 f(x; \theta) \partial_x^2 f(x; \theta)^T + \tfrac{1}{2}(\partial_x^3 f(x; \theta) \partial_x f(x; \theta)^T + \partial_x f(x; \theta) \partial_x^3 f(x; \theta)^T)\right)\right] + O(\Delta^3), \end{aligned}$$

where

$$\begin{aligned} q_1(x;\theta) &= \tfrac{1}{2}[b(x;\alpha)(2+\partial_x v(x;\beta)) - 2b(x;\alpha) + \tfrac{1}{2}v(x;\beta)(4\partial_x b(x;\alpha) + \partial_x^2 v(x;\beta))] \\ q_2(x,\theta) &= \tfrac{3}{4}v(x;\beta)(1+\tfrac{1}{3}b(x;\alpha)+\partial_x v(x;\beta)). \end{aligned}$$

Since

$$\begin{aligned} \partial_\alpha L_\theta f(x;\theta) - L_\theta \partial_\alpha f(x;\theta) &= \partial_\alpha b(x;\alpha)\partial_x f(x;\theta) \\ \partial_\beta L_\theta f(x;\theta) - L_\theta \partial_\beta f(x;\theta) &= \tfrac{1}{2}\partial_\beta v(x;\beta)\partial_x^2 f(x;\theta) \end{aligned}$$

it also follows from (5.19) that

$$\partial_{\theta^T}\pi_\theta^\Delta f(x) - \pi_\theta^\Delta \partial_{\theta^T} f(x) = \Delta F(x) \begin{pmatrix} \partial_\alpha b(x;\alpha) & 0 \\ 0 & \tfrac{1}{2}\partial_\beta v(x;\beta) \end{pmatrix} + O(\Delta^2),$$

where $F(x)$ denotes the $N \times 2$-matrix $F(x) = (\partial_x f(x), \partial_x^2 f(x))$.

If $A^*(x,\Delta;\theta)$ satisfies (4.8), then the $2 \times N$-matrix

$$B(x,\Delta;\theta) = \begin{pmatrix} 1 & 0 \\ 0 & 2\Delta \end{pmatrix} A^*(x,\Delta;\theta).$$

satisfies that

$$\begin{aligned} &B(x,\Delta;\theta)\left[v(x;\beta)\partial_x f(x;\theta)\partial_x f(x;\theta)^T + \Delta M(x;\theta) + O(\Delta^2)\right] \\ &\qquad = \begin{pmatrix} \partial_\alpha b(x;\alpha) & 0 \\ 0 & \Delta\partial_\beta v(x;\beta) \end{pmatrix} F(x)^T + \begin{pmatrix} O(\Delta) \\ O(\Delta^2) \end{pmatrix} \end{aligned}$$

where $\Delta^2 M(x;\theta)$ denotes the term of order Δ^2 in (5.20). Let $B(x,\Delta;\theta)_i$ denote the ith row of $B(x,\Delta;\theta)$ $(i=1,2)$. Then it follows by letting Δ tend to zero that

$$v(x;\beta)B(x,0;\theta)_2\partial_x f(x;\theta)\partial_x f(x;\theta)^T = 0. \tag{5.21}$$

The condition that $D(x)$ is invertible implies that we can find a coordinate of $\partial_x f(x;\theta)$ which is not equal to zero, so we conclude that

$$\partial_y g_2^*(0,x,x;\theta) = B(x,0;\theta)_2\partial_x f(x;\theta) = 0.$$

Similarly we find that

$$[v(x;\beta)B(x,0;\theta)_1\partial_x f(x;\theta) - \partial_\alpha b(x;\alpha)]\partial_x f(x;\theta)^T = 0,$$

which implies

$$\partial_y g_1^*(0,x,x;\theta) = B(x,0;\theta)_1\partial_x f(x;\theta) = \partial_\alpha b(x;\alpha)/v(x;\beta).$$

Finally, (5.21) implies that

$$B(x,0;\theta)_2 M(x;\theta) = \partial_\beta v(x;\beta)\partial_x^2 f(x;\theta)^T.$$

Since we have shown that $B(x,0;\theta)_2\partial_x f(x;\theta) = 0$, this expression simplifies to

$$\begin{aligned}[q_2(x;\theta)B(x,0;\theta)_2\partial_x^2 f(x;\theta) + \tfrac{1}{2}v(x;\beta)^2 B(x,0;\theta)_2\partial_x^3 f(x;\theta)]\partial_x f(x;\theta)^T \\ = [\partial_\beta v(x;\beta) - v(x;\beta)^2 B(x,0;\theta)_2\partial_x^2 f(x;\theta)]\partial_x^2 f(x;\theta)^T.\end{aligned}$$

Thus real functions $c_1(x;\theta)$ and $c_2(x;\theta)$ exist such that $c_1(x;\theta)\partial_x f(x;\theta) = c_2(x;\theta)\partial_x^2 f(x;\theta)$. If $c_2(x;\theta) \neq 0$, then $\partial_x^2 f(x) = c_1(x;\theta)/c_2(x;\theta)\partial_x f(x;\theta)$, which implies that $\det(D(x)) = 0$. Thus we can conclude that $\partial_\beta v(x;\beta) - v(x;\beta)^2 B(x,0;\theta)_2\partial_x^2 f(x;\theta) = c_2(x;\theta) = 0$ or

$$\partial_y^2 g_2^*(0,x,x;\theta) = B(x,0;\theta)_2\partial_x^2 f(x;\theta) = \partial_\beta v(x;\beta)/v(x;\beta)^2.$$

□

6 Conclusions

A general theory of high frequency asymptotics has been developed for a large class of estimators, essentially any estimator that can be obtained from estimating functions or the generalized method of moments based on conditional moments or on approximations to conditional moments. Simple conditions have been derived that ensure rate optimality and efficiency of the estimators. For diffusion models it is important to use rate optimal estimators, because otherwise the information about the diffusion coefficient contained in the quadratic variation is not used. A number of previously proposed estimators have been shown to satisfy the conditions for rate optimality and efficiency, including the maximum likelihood estimator, the estimator based on the Gaussian Euler approximation to the likelihood function, other similar maximum pseudo-likelihood estimators, and Godambe-Heyde optimal martingale estimating functions. Tools for studying high frequency asymptotic properties of estimators have been provided, including in particular simple conditions ensuring that convergence in probability of a normalized sum of parameter-dependent functions of pairs of consecutive observations is uniform in the parameter.

Acknowledgements

I am grateful to Arnaud Gloter, Mathieu Kessler, Masayuki Uchida and Nakahiro Yoshida for helpful discussions. Part of the paper was written during a stay at the Mittag-Leffler Institute. The research was supported by the Danish Center for Accounting and Finance funded by the Danish Social Science Research Council and by the Center for Research in Econometric Analysis of Time Series funded by the Danish National Research Foundation.

References

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

Aït-Sahalia, Y. & Mykland, P. (2003). "The effects of random and discrete sampling when estimating continuous-time diffusions". *Econometrica*, 71:483–549.

Aït-Sahalia, Y. & Mykland, P. A. (2004). "Estimators of diffusions with randomly spaced discrete observations: a general theory". *Ann. Statist.*, 32:2186–2222.

Barndorff-Nielsen, O. E. & Sørensen, M. (1994). "A review of some aspects of asymptotic likelihood theory for stochastic processes". *International Statistical Review*, 62:133–165.

Beskos, A.; Papaspiliopoulos, O.; Roberts, G. O. & Fearnhead, P. (2006). "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes". *J. Roy. Statist. Soc. B*, 68:333–382.

Bibby, B. M.; Jacobsen, M. & Sørensen, M. (2010). "Estimating functions for discretely sampled diffusion-type models". In Aït-Sahalia, Y. & Hansen, L. P., editors, *Handbook of Financial Econometrics*, pages 203–268. North Holland, Oxford.

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

Bibby, B. M. & Sørensen, M. (1996). "On estimation for discretely observed diffusions: A Review". *Theory of Stochastic Processes*, 2:49–56.

Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

Campbell, J. Y.; Lo, A. W. & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

Chan, K. C.; Karolyi, G. A.; Longstaff, F. A. & Sanders, A. B. (1992). "An empirical comparison of alternative models of the short-term interest rate". *Journal of Finance*, 47:1209–1227.

Christensen, B. J. & Sørensen, M. (2008). "Optimal inference in dynamic models with conditional moment restrictions". Preprint, Creates.

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

Ditlevsen, P.; Ditlevsen, S. & Andersen, K. K. (2002). "The fast climate fluctuations during the stadial and interstadial climate states". *Annals of Glaciology*, 35:457–462.

Ditlevsen, S.; Yip, K.-P.; Marsh, D. J. & Holstein-Rathlou, N.-H. (2007). "Parameter estimation of the feedback gain in a stochastic model of renal hemodynamics". *American Journal of Physiology. Renal Physiology*, 292:607–616.

Durham, G. B. & Gallant, A. R. (2002). "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes". *J. Business & Econom. Statist.*, 20:297–338.

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

Genon-Catalot, V. & Jacod, J. (1993). "On the estimation of the diffusion coefficient for multi-dimensional diffusion processes". *Ann. I. H. Poincaré PR*, 29:119–151.

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

Gihman, I. I. & Skorohod, A. V. (1972). *Stochastic Differential Equations.* Springer-Verlag, Berlin.

Gloter, A. & Sørensen, M. (2009). "Estimation for stochastic differential equations with a small diffusion coefficient". *Stoch. Proc. Appl.*, 119:679–699.

Gobet, E. (2002). "LAN property for ergodic diffusions with discrete observations". *Ann. I. H. Poincaré PR*, 38:711–737.

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

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

Hansen, L. P. (1982). "Large sample properties of generalized method of moments estimators". *Econometrica*, 50:1029–1054.

Hansen, L. P. (1985). "A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators". *Journal of Econometrics*, 30:203–238.

Hansen, L. P. (1993). "Semiparametric efficiency bounds for linear time-series models". In Phillips, P. C. B., editor, *Models, Methods and Applications of Econometrics: Esseys in Honor of A. R. Bergstrom*, pages 253–271. Cambridge, MA: Blackwell.

Hansen, L. P.; Scheinkman, J. A. & Touzi, N. (1998). "Spectral methods for identifying scalar diffusions". *Journal of Econometrics*, 86:1–32.

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

Ibragimov, I. A. & Has'minskii, R. Z. (1981). *Statistical Estimation - Asymptotic Theory.* Springer-Verlag, New York.

Jacobsen, M. (2001). "Discretely observed diffusions; classes of estimating functions and small Δ-optimality". *Scand. J. Statist.*, 28:123–150.

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

Jacod, J. & Sørensen, M. (2010). "Asymptotic statistical theory for stochastic processes: A review". Preprint, Department of Mathematical Sciences, University of Copenhagen. In preparation.

Kallenberg, O. (1997). *Foundations of Modern Probability.* Springer-Verlag, New York.

Kelly, L.; Platen, E. & Sørensen, M. (2004). "Estimation for discretely observed diffusions using transform functions". *J. Appl. Prob.*, 41:99–118.

Kessler, M. (1997). "Estimation of an ergodic diffusion from discrete observations". *Scand. J. Statist.*, 24:211–229.

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

Kutoyants, Y. (2004). *Statistical Inference for Ergodic Diffusion Processes.* Springer, New York.

Lansky, P.; Sacerdote, L. & Tomasetti, F. (1995). "On comparison of Feller and Ornstein-Uhlenbeck models for neural activity". *Biol. Cybern.*, 73:457–465.

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

McAdams, H. & Arkin, A. (1997). "Stochastic mechanisms in gene expression". *Proc. Nat. Acad. Sci. USA*, 94:814–819.

Overbeck, L. & Rydén, T. (1997). "Estimation in the Cox-Ingersoll-Ross model". *Econometric Theory*, 13:430–461.

Pedersen, A. R. (1995). "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations". *Scand. J. Statist.*, 22:55–71.

Pedersen, A. R. (2000). "Estimating the nitrous oxide emission rate from the soil surface by means of a diffusion model". *Scand. J. Statist.*, 27:385–403.

Pokern, Y.; Stuart, A. M. & Wiberg, P. (2009). "Parameter estimation for partially observed hypoelliptic diffusions". *J. Roy. Statist. Soc. B*, 71:49–73.

Poulsen, R. (1999). "Approximate maximum likelihood estimation of discretely observed diffusion processes". Working Paper 29, Centre for Analytical Finance, Aarhus.

Prakasa Rao, B. L. S. (1988). "Statistical inference from sampled data for stochastic processes". *Contemporary Mathematics*, 80:249–284.

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

Sørensen, M. & Uchida, M. (2003). "Small-diffusion asymptotics for discretely sampled stochastic differential equations". *Bernoulli*, 9:1051–1069.

Yoshida, N. (1990). "Asymptotic behaviour of M-estimator and related random field for diffusion process". *Ann. Inst. Statist. Math*, 42:221–251.

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