Electronic Journal of Statistics
Vol. 15 (2021) 1424–1472
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1777

Adaptive estimation for degenerate diffusion processes*

Arnaud Gloter

Université Paris-Saclay, CNRS, Univ Evry,
Laboratoire de Mathématiques et Modélisation d'Evry,
91037, Evry-Courcouronnes, France.
e-mail: arnaud.gloter@univ-evry.fr

and

Nakahiro Yoshida

Graduate School of Mathematical Sciences, University of Tokyo
3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan.
e-mail: nakahiro@ms.u-tokyo.ac.jp
url: https://www.ms.u-tokyo.ac.jp/~nakahiro/hp-naka-e

Abstract: We discuss parametric estimation of a degenerate diffusion system from time-discrete observations. The first component of the degenerate diffusion system has a parameter θ_1 in a non-degenerate diffusion coefficient and a parameter θ_2 in the drift term. The second component has a drift term parameterized by θ_3 and no diffusion term. Asymptotic normality is proved in two different situations for an adaptive estimator for θ_3 with some initial estimators for (θ_1, θ_2), and an adaptive one-step estimator for $(\theta_1, \theta_2, \theta_3)$ with some initial estimators for them. Our estimators incorporate information of the increments of both components. Thanks to this construction, the asymptotic variance of the estimators for θ_1 is smaller than the standard one based only on the first component. The convergence of the estimators for θ_3 is much faster than the other parameters. The resulting asymptotic variance is smaller than that of an estimator only using the increments of the second component.

Keywords and phrases: Degenerate diffusion, one-step estimator, quasi-maximum likelihood estimator.

Received February 2020.

Contents

*This work was in part supported by Japan Science and Technology Agency CREST JPMJCR14D7; Japan Society for the Promotion of Science Grants-in-Aid for Scientific Research No. 17H01702 (Scientific Research); and by a Cooperative Research Program of the Institute of Statistical Mathematics.

1. Introduction

In this article, we will discuss parametric estimation for a hypo-elliptic diffusion process. More precisely, given a stochastic basis $(\Omega, \mathcal{F}, \mathbf{F}, P)$ with a right-continuous filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathbb{R}_+}$, $\mathbb{R}_+ = [0,\infty)$, suppose that an $\mathbf{F}$-adapted process $Z_t = (X_t, Y_t)$ satisfies the stochastic differential equation

$$\begin{cases} dX_t &= A(Z_t,\theta_2)dt + B(Z_t,\theta_1)dw_t \\ dY_t &= H(Z_t,\theta_3)dt. \end{cases} \tag{1.1}$$

Here $A : \mathbb{R}^{\mathsf{d}_Z} \times \overline{\Theta}_2 \to \mathbb{R}^{\mathsf{d}_X}$, $B : \mathbb{R}^{\mathsf{d}_Z} \times \overline{\Theta}_1 \to \mathbb{R}^{\mathsf{d}_X} \otimes \mathbb{R}^{\mathsf{r}}$, $H : \mathbb{R}^{\mathsf{d}_Z} \times \overline{\Theta}_3 \to \mathbb{R}^{\mathsf{d}_Y}$, and $w = (w_t)_{t\in\mathbb{R}_+}$ is an r-dimensional $\mathbf{F}$-Wiener process. The spaces Θ_i $(i = 1,2,3)$ are the unknown parameter spaces of the components of $\theta = (\theta_1, \theta_2, \theta_3)$ to be estimated from the data $(Z_{t_j})_{j=0,1,\dots,n}$, where $t_j = jh$, $h = h_n$ satisfying $h \to 0$, $nh \to \infty$ and $nh^2 \to 0$ as $n \to \infty$.

Estimation theory has been well developed for diffusion processes. Even focusing on parametric estimation for ergodic diffusions, there is huge amount of studies: Kutoyants [27, 29, 28], Prakasa Rao [39, 40], Yoshida [50, 51, 52], Bibby and Sørensen [2], Kessler [24], Küchler and Sørensen [25], Genon–Catalot et al. [14], Gloter [16, 17, 19], Sakamoto and Yoshida [41], Uchida [45], Uchida and Yoshida [46, 47, 48], Kamatani and Uchida [23], De Gregorio and Iacus [10], Genon–Catalot and Larédo [15], Suzuki and Yoshida [44] among many others. Nakakita and Uchida [36] and Nakakita et al. [35] studied estimation under measurement error; related are Gloter and Jacod [20, 21]. Non parametric estimation for the coefficients of an ergodic diffusion has also been widely studied: Dalayan and Kutoyants [9], Kutoyants [29], Dalalyan [6], Dalalyan and Reiss [7, 8], Comte and Genon–Catalot [3], Comte et al. [4], Schmisser [43] to name a few. Historically attentions were paid to inference for non-degenerate cases.

Recently there is a growing interest in hypo-elliptic diffusions, that appear in various applied fields. Examples of the hypo-elliptic diffusion include the harmonic oscillator, the Van der Pol oscillator and the FitzHugh-Nagumo neuronal model; see e.g. León and Samson [30]. For parametric estimation of hypo-elliptic diffusions, we refer the reader to Gloter [18] for a discretely observed integrated diffusion process, and Samson and Thieullen [42] for a contrast estimator. Comte et al. [5] gave adaptive estimation under partial observation. Recently, Ditlevsen and Samson [12] studied filtering and inference for hypo-elliptic diffusions from complete and partial observations. When the observations are discrete and complete, they showed asymptotic normality of their estimators under the assumption that the true value of some of parameters are known. Melnykova [33] studied the estimation problem for the model (1.1), comparing contrast functions and least square estimates. The contrast functions we propose in this paper are different from the one in [33]. Recently, Delattre et al. [11] gave a rate of convergence to a nonparametric estimator for the stationary distribution of a hypoelliptic diffusion.

In this paper, we will present several estimation schemes. Since we assume discrete-time observations of $Z= (Z_t)_{t\in\mathbb{R}+}$, quasi-likelihood estimation for θ_1 and θ_2 is known; only difference from the standard diffusion case is the existence of the covariate $Y= (Y_t)_{t\in\mathbb{R}+}$ in the equation of $X= (X_t)_{t\in\mathbb{R}+}$ but it causes no theoretical difficulty. Thus, our first approach in Section 3 is toward estimation of θ_3 with initial estimators for θ_1 and θ_2. The idea for construction of the quasi-likelihood function in the elliptic case was based on the local Gaussian approximation of the transition density. Then it is natural to approximate the distribution of the increments of Y by that of the principal Gaussian variable in the expansion of the increment. However, this method causes deficiency, as we will observe there; see Section 8. We present a more efficient method by incorporating an additional Gaussian part from X. The error rate attained by the estimator for θ_3 is $n^{-1/2}h^{1/2}$ and it is much faster than the rate $(nh)^{-1/2}$ for θ_2 and $n^{-1/2}$ for θ_1. Section 4 treats some adaptive estimators using suitable initial estimators for $(\theta_1, \theta_2, \theta_3)$, and shows joint asymptotic normality. Then it should be remarked that the asymptotic variance of our estimator $\hat{\theta}_1$ for θ_1 has improved that of the ordinary volatility parameter estimator, e.g. $\hat{\theta}_1^0$ recalled in Section 3.4 that would be asymptotically optimal if the system consisted only of X. Section 2 collects the assumptions under which we will work. Section 5 offers several basic estimates to the increments of Z.

To investigate efficiency of the presented estimators, we need the LAN property of the exact likelihood function of the hypo-elliptic diffusion. Another important and natural question the reader must have is the asymptotic behavior of the joint quasi-maximum likelihood estimator based on a quasi-likelihood random field for the full parameter θ; an expression of the random field has already appeared in (4.2) essentially. In the present situation, the three parameters have different convergence rates and in particular the handling of θ_3 is not straightforward because for estimation of θ_3, the parameters (θ_1, θ_2) become nuisance, but any estimator of them has very large error compared with θ_3. The user could get some estimated value with the joint quasi-likelihood random field, however, there is no theoretical backing for such a scheme. Though somewhat sophisticated treatments are necessary, we can validate the joint quasi-maximum likelihood estimator and can show that the same asymptotic variance is attained, up to the first order, as the one-step quasi-likelihood estimator provided in this article. We will discuss these problems elsewhere, while we recommend the reader to see Gloter and Yoshida [22] for more complete exposition including the non-adaptive approach and additional information.

2. Assumptions

We assume that Θ_i $(i = 1, 2, 3)$ are bounded open domain in $\mathbb{R}^{\mathsf{p}_i}$, respectively, and $\Theta = \prod_{i=1}^{3}\Theta_i$ has a good boundary so that Sobolev's embedding inequality (cf. Adams [1]) holds, that is, there exists a positive constant C_Θ such that

$$\sup_{\theta\in\Theta}|f(\theta)| \leq C_\Theta \sum_{k=0}^{1} \|\partial_\theta^k f\|_{L^p(\Theta)} \tag{2.1}$$

for all $f \in C^1(\Theta)$ and $p > \sum_{i=1}^{3} \mathsf{p}_i$. If Θ has a Lipschitz boundary, then this condition is satisfied. Obviously, the embedding inequality (2.1) is valid for functions depending only on a part of components of θ. In this paper, giving priority to simplicity of presentation, we use Sobolev's inequality to control the maximum of a random field though other embedding inequalities such as the GRR inequality improve the assumptions on differentiability of the coefficients of the stochastic differential equations.

In this paper, we will propose an estimator for θ and show its consistency and asymptotic normality.

Given a finite-dimensional real vector space E, denote by $C_p^{a,b}(\mathbb{R}^{\mathsf{d}_Z} \times \Theta_i; \mathsf{E})$ the set of functions $f : \mathbb{R}^{\mathsf{d}_Z} \times \Theta_i \to \mathsf{E}$ such that f is continuously differentiable a times in $z \in \mathbb{R}^{\mathsf{d}_Z}$ and b times in $\theta_i \in \Theta_i$ in any order and f and all such derivatives are continuously extended to $\mathbb{R}^{\mathsf{d}_Z} \times \overline{\Theta}_i$, moreover, they are of at most polynomial growth in $z \in \mathbb{R}^{\mathsf{d}_Z}$ uniformly in $\theta_i \in \Theta_i$. Let

$$C = BB^\star,$$

where $\star$ denotes the matrix transpose. We suppose that the process $(Z_t)_{t\in\mathbb{R}_+}$ generating the data satisfies the stochastic differential equation (1.1) for a true value $\theta^* = (\theta_1^*, \theta_2^*, \theta_3^*) \in \Theta_1 \times \Theta_2 \times \Theta_3$.

[**A1**] **(i)** $A \in C_p^{i_A,j_A}(\mathbb{R}^{\mathsf{d}_Z} \times \Theta_2; \mathbb{R}^{\mathsf{d}_X})$ and $B \in C_p^{i_B,j_B}(\mathbb{R}^{\mathsf{d}_Z} \times \Theta_1; \mathbb{R}^{\mathsf{d}_X} \otimes \mathbb{R}^{\mathsf{r}})$.

(ii) $H \in C_p^{i_H,j_H}(\mathbb{R}^{\mathsf{d}_Z} \times \Theta_3; \mathbb{R}^{\mathsf{d}_Y})$.

We will denote F_x for $\partial_x F$, F_y for $\partial_y F$, and F_i for $\partial_{\theta_i} F$.

[**A2**] **(i)** $\sup_{t\in\mathbb{R}_+}\|Z_t\|_p < \infty$ for every $p > 1$.

(ii) There exists a probability measure ν on $\mathbb{R}^{\mathsf{d}_Z}$ such that

$$\frac{1}{T}\int_0^T f(Z_t)\,dt \to^p \int f(z)\nu(dz) \qquad (T \to \infty)$$

for any bounded continuous function $f : \mathbb{R}^{\mathsf{d}_Z} \to \mathbb{R}$.

(iii) The function $\theta_1 \mapsto C(Z_t, \theta_1)^{-1}$ is continuous on $\overline{\Theta}_1$ a.s., and

$$\sup_{\theta_1 \in \Theta_1} \sup_{t \in \mathbb{R}_+} \| \det C(Z_t, \theta_1)^{-1} \|_p < \infty$$

for every $p > 1$.

(iv) For the $\mathbb{R}^{\mathsf{d}_Y} \otimes \mathbb{R}^{\mathsf{d}_Y}$ valued function

$$V(z, \theta_1, \theta_3) = H_x(z, \theta_3) C(z, \theta_1) H_x(z, \theta_3)^\star, \tag{2.2}$$

the function $(\theta_1, \theta_3) \mapsto V(Z_t, \theta_1, \theta_3)^{-1}$ is continuous on $\overline{\Theta}_1 \times \overline{\Theta}_3$ a.s., and

$$\sup_{(\theta_1, \theta_3) \in \Theta_1 \times \Theta_3} \sup_{t \in \mathbb{R}_+} \| \det V(Z_t, \theta_1, \theta_3)^{-1} \|_p < \infty$$

for every $p > 1$.

Remark 2.1. **(a)** It follows from $[A2]$ that the convergence in $[A2]$ (ii) holds for any continuous function f of at most polynomial growth.

(b) We implicitly assume the existence of $C(Z_t, \theta_1)^{-1}$ and $V(Z_t, \theta_1, \theta_3)^{-1}$ in (iii) and (iv) of $[A2]$.

(c) Fatou's lemma implies

$$\int |z|^p \nu(dz) + \sup_{\theta_1 \in \overline{\Theta}_1} \int \big(\det C(z, \theta_1) \big)^{-p} \nu(dz) + \sup_{(\theta_1, \theta_3) \in \overline{\Theta}_1 \times \overline{\Theta}_3} \int \big(\det V(z, \theta_1, \theta_3) \big)^{-p} \nu(dz) < \infty$$

for any $p > 0$.

(d) Assumption $[A2]$ is standard. Exponential ergodicity and boundedness of any order of moment of the process are also well known. For nondegenerate diffusions, see e.g. Pardoux and Veretennikov [38], Meyn and Tweedie [34] and Kusuoka and Yoshida [26] among many others. For damping Hamiltonian systems, we refer the reader to Wu [49]. The Lyapounov function method provides exponential mixing (even in the non-stationary case) and estimates of moments of the invariant probability measure up to any order. Wu's paper investigated several examples including the van der Pol model. We are giving additional information in Delattre et al. [11].

Let

$$\mathbb{Y}^{(1)}(\theta_1) = -\frac{1}{2} \int \left\{ \mathrm{Tr}\big(C(z, \theta_1)^{-1} C(z, \theta_1^*) \big) - \mathsf{d}_X + \log \frac{\det C(z, \theta_1)}{\det C(z, \theta_1^*)} \right\} \nu(dz).$$

Since $|\log x| \le x + x^{-1}$ for $x > 0$, $\mathbb{Y}^{(1)}(\theta_1)$ is a continuous function on $\overline{\Theta}_1$ well defined under $[A1]$ and $[A2]$. Let

$$\mathbb{Y}^{(2)}(\theta_2) = -\frac{1}{2} \int C(z, \theta_1^*)^{-1} \big[\big(A(z, \theta_2) - A(z, \theta_2^*) \big)^{\otimes 2} \big] \nu(dz), \tag{2.3}$$

and

$$\mathbb{Y}^{(3)}(\theta_3) = -\int 6V(z,\theta_1^*,\theta_3)^{-1}\big[\big(H(z,\theta_3)-H(z,\theta_3^*)\big)^{\otimes 2}\big]\nu(dz).$$

The random field $\mathbb{Y}^{(3)}$ is well defined under $[A1]$ and $[A2]$. Obviously, ν depends on the value θ^*. We suppress θ^* from notation since it is fixed in this article, where it is not necessary to change θ^* differently from discussion of the asymptotic minimax bound for example.

We will assume all or some of the following identifiability conditions

[A3] **(i)** There exists a positive constant χ_1 such that

$$\mathbb{Y}^{(1)}(\theta_1) \le -\chi_1|\theta_1-\theta_1^*|^2 \qquad (\theta_1\in\Theta_1).$$

(ii) There exists a positive constant χ_2 such that

$$\mathbb{Y}^{(2)}(\theta_2) \le -\chi_2|\theta_2-\theta_2^*|^2 \qquad (\theta_2\in\Theta_2).$$

(iii) There exists a positive constant χ_3 such that

$$\mathbb{Y}^{(3)}(\theta_3) \le -\chi_3|\theta_3-\theta_3^*|^2 \qquad (\theta_3\in\Theta_3).$$

In the hypoelliptic case, as it is the most interesting case, checking these identifiability conditions is usually easy since ν is equivalent to or at least dominated by the Lebesgue measure and admits a density that is positive on a non-empty open set. Thus, identifiability is a problem of parameterization of the model. In particular, it is obvious that this condition causes no difficulty for linearly parametrized models often appearing in applications.

As already mentioned, we will assume that $h\to 0$, $nh\to\infty$ and $nh^2\to 0$ as $n\to\infty$ throughout this article. The condition $nh^2\to 0$ is a standard one called the condition for rapidly increasing experimental design (Prakasa Rao [40]). Yoshida [51] relaxed this condition to $nh^3\to 0$, and Kessler [24] to $nh^p\to 0$ for any positive number p. Uchida and Yoshida [48] carried out the Ibragimov-Has'minskii-Kutoyants program under the condition $nh^p\to 0$, with the so-called Quasi-Likelihood Analysis based on the polynomial type large deviation estimate for the quasi-likelihood random field (Yoshida [52]). It is well known that these approaches under $nh^p\to 0$ need more smoothness of the model than our assumptions because they inevitably involve higher-order expansions of the semigroup. In this paper, when estimating the order of a random variable, eventually we use either $n\to\infty$, $h\to 0$, $nh\to\infty$ or $nh^2\to 0$, and that's all. So, it is easy for the reader to recognize which convergence is used in each case. For example, if the reader finds $O_p(\sqrt{n}h)$, then quite likely it will be estimated as $o_p(1)$. However, we left traces as many as possible in the proof.

3. Adaptive estimation of θ_3

We denote $U^{\otimes k}$ for $U\otimes\cdots\otimes U$ (k-times) for a tensor U. For tensors $S^1 = (S^1_{i_{1,1},\dots,i_{1,d_1};j_{1,1},\dots,j_{1,k_1}})$, ..., $S^m = (S^m_{i_{m,1},\dots,i_{m,d_m};j_{m,1},\dots,j_{m,k_m}})$ and and a tensor

$T = (T^{i_{1,1},\dots,i_{1,d_1},\dots,i_{m,1},\dots,i_{m,d_m}})$, we write

$$\begin{aligned} & T[S^1, ..., S^m] \;=\; T[S^1 \otimes \cdots \otimes S^m] \\ = \; & \Bigg(\sum_{i_{1,1},\dots,i_{1,d_1},\dots,i_{m,1},\dots,i_{m,d_m}} T^{i_{1,1},\dots,i_{1,d_1},\dots,i_{m,1},\dots,i_{m,d_m}} S^1_{i_{1,1},\dots,i_{1,d_1};j_{1,1},\dots,j_{1,k_1}} \\ & \qquad \cdots S^m_{i_{m,1},\dots,i_{m,d_m};j_{m,1},\dots,j_{m,k_m}} \Bigg)_{j_{1,1},\dots,j_{1,k_1},\dots,j_{m,1},\dots,j_{m,k_m}}. \end{aligned}$$

This notation will be applied for a tensor-valued tensor T as well.

Remark 3.1. Clearly, this notation has an advantage over the notation by matrix product since the elements $S^1, ..., S^m$ quite often have a long expression in the inference. The matrix notation repeats S^is twice for the quadratic form, thrice for the cubic form, and so on. This notation was introduced by [52] and already adopted by many papers, e.g., [45], [48], [53], [31], [23], [32], [13], [37], [36], [35], just to name a few.

Let

$$\begin{aligned} L_H(z, \theta_1, \theta_2, \theta_3) \;\; &= \;\; H_x(z, \theta_3)[A(z, \theta_2)] + \frac{1}{2} H_{xx}(z, \theta_3)[C(z, \theta_1)] \\ & \quad\; + H_y(z, \theta_3)[H(z, \theta_3)]. \end{aligned}$$

Define the $\mathbb{R}^{\mathsf{d}_Y}$-valued function $G_n(z, \theta_1, \theta_2, \theta_3)$ by

$$G_n(z, \theta_1, \theta_2, \theta_3) = H(z, \theta_3) + \frac{h}{2} L_H\big(z, \theta_1, \theta_2, \theta_3\big). \tag{3.1}$$

Let

$$\mathcal{D}_j(\theta_1, \theta_2, \theta_3) = \begin{pmatrix} h^{-1/2}\big(\Delta_j X - hA(Z_{t_{j-1}}, \theta_2)\big) \\ h^{-3/2}\big(\Delta_j Y - hG_n(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3)\big) \end{pmatrix}. \tag{3.2}$$

We will work with some initial estimators $\hat{\theta}^0_1$ for θ^0_1 and $\hat{\theta}^0_2$ for θ_2. The following standard convergence rates, in part or fully, will be assumed for these estimators:

[**A4**] **(i)** $\hat{\theta}^0_1 - \theta^*_1 \;=\; O_p(n^{-1/2})$ as $n \to \infty$

(ii) $\hat{\theta}^0_2 - \theta^*_2 \;=\; O_p(n^{-1/2}h^{-1/2})$ as $n \to \infty$

The expansions (5.1) and (5.6) with Lemma 5.5 suggest two approaches for estimating θ_3. The first approach is based on the likelihood of $\Delta_j Y$ only, without assistance of $\Delta_j X$. The second one uses the likelihood corresponding to $\mathcal{D}_j(\theta_1, \theta_2, \theta_3)$. However, it is possible to show that the first approach gives less optimal asymptotic variance; see Section 8. So, we will take the second approach here.

3.1. Adaptive quasi-likelihood function for θ_3

Recall (2.2):

$$V(z,\theta_1,\theta_3) = H_x(z,\theta_3)C(z,\theta_1)H_x(z,\theta_3)^\star.$$

Let

$$S(z,\theta_1,\theta_3) = \begin{pmatrix} C(z,\theta_1) & 2^{-1}C(z,\theta_1)H_x(z,\theta_3)^\star \\ 2^{-1}H_x(z,\theta_3)C(z,\theta_1) & 3^{-1}H_x(z,\theta_3)C(z,\theta_1)H_x(z,\theta_3)^\star \end{pmatrix}.$$

Then

$$S(z,\theta_1,\theta_3)^{-1} = \begin{pmatrix} S(z,\theta_1,\theta_3)^{1,1} & S(z,\theta_1,\theta_3)^{1,2} \\ S(z,\theta_1,\theta_3)^{2,1} & S(z,\theta_1,\theta_3)^{2,2} \end{pmatrix}, \tag{3.3}$$

where

$$\begin{aligned} S(z,\theta_1,\theta_3)^{1,1} &= C(z,\theta_1)^{-1} + 3H_x(z,\theta_3)^\star V(z,\theta_1,\theta_3)^{-1}H_x(z,\theta_3), \\ S(z,\theta_1,\theta_3)^{1,2} &= -6H_x(z,\theta_3)^\star V(z,\theta_1,\theta_3)^{-1}, \\ S(z,\theta_1,\theta_3)^{2,1} &= -6V(z,\theta_1,\theta_3)^{-1}H_x(z,\theta_3) \end{aligned}$$

and

$$S(z,\theta_1,\theta_3)^{2,2} = 12V(z,\theta_1,\theta_3)^{-1}.$$

Let

$$\hat{S}(z,\theta_3) = S(z,\hat{\theta}_1^0,\theta_3).$$

Since the increment $\Delta_j Z = Z_{t_j} - Z_{t_{j-1}}$ is approximately conditionally Gaussian in short-term asymptotics, it seems natural to construct a likelihood function based on the local Gaussian approximation. Remark that

$$S(z,\theta_1,\theta_3) = \begin{pmatrix} B(z,\theta_1)B(z,\theta_1)^\star & \frac{\sqrt{3}}{2}B(z,\theta_1)\kappa(z,\theta_1,\theta_3)^\star \\ \frac{\sqrt{3}}{2}\kappa(z,\theta_1,\theta_3)B(z,\theta_1)^\star & \kappa(z,\theta_1,\theta_3)\kappa(z,\theta_1,\theta_3)^\star \end{pmatrix}$$

is the covariance matrix of the principal conditionally Gaussian part of $(\Delta_j X, \Delta_j Y)$ if properly scaled and evaluated at $z = Z_{t_{j-1}}$ and $(\theta_1,\theta_3) = (\theta_1^*,\theta_3^*)$, where

$$\kappa(z,\theta_1,\theta_3) = 3^{-1/2}H_x(z,\theta_3)B(z,\theta_1). \tag{3.4}$$

See Lemmas 5.4 and 5.5.

We define a log quasi-likelihood function by

$$\mathbb{H}_n^{(3)}(\theta_3) \quad = \quad -\frac{1}{2}\sum_{j=1}^n \left\{ \hat{S}(Z_{t_{j-1}},\theta_3)^{-1}\left[\mathcal{D}_j(\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)^{\otimes 2}\right] + \log\det \hat{S}(Z_{t_{j-1}},\theta_3) \right\}. \tag{3.5}$$

Let $\hat{\theta}_3^0$ be a quasi-maximum likelihood estimator (QMLE) for θ_3 for $\mathbb{H}_n^{(3)}$, that is, $\hat{\theta}_3^0$ is a $\overline{\Theta}_3$-valued measurable mapping satisfying

$$\mathbb{H}_n^{(3)}(\hat{\theta}_3^0) = \max_{\theta_3 \in \overline{\Theta}_3} \mathbb{H}_n^{(3)}(\theta_3).$$

The QMLE $\hat{\theta}_3^0$ for $\mathbb{H}_n^{(3)}$ depends on n as it does on the data $(Z_{t_j})_{j=0,1,\dots,n}$; $\hat{\theta}_1^0$ in the function $\hat{S}$ also depends on $(Z_{t_j})_{j=0,1,\dots,n}$.

3.2. Consistency of $\hat{\theta}_3^0$

Let

$$\mathbb{Y}_n^{(3)}(\theta_3) = n^{-1}h\big\{\mathbb{H}_n^{(3)}(\theta_3) - \mathbb{H}_n^{(3)}(\theta_3^*)\big\}.$$

Theorem 3.2. *Suppose that* [A1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,3,1)$ *and* [A2] *are satisfied. Then*

$$\sup_{\theta_3\in\overline{\Theta}_3} \left|\mathbb{Y}_n^{(3)}(\theta_3) - \mathbb{Y}^{(3)}(\theta_3)\right| \to^p 0 \tag{3.6}$$

as $n\to\infty$, *if* $\hat{\theta}_1^0 \to^p \theta_1^*$ *and* $\hat{\theta}_2^0 \to^p \theta_2^*$. *Moreover,* $\hat{\theta}_3^0 \to^p \theta_3^*$ *if* [A3] *(iii) is additionally satisfied.*

Proof of Theorem 3.2 is in Section 6.

3.3. Asymptotic normality of $\hat{\theta}_3^0$

Let

$$\begin{aligned}\Gamma_{33} &= \int S(z,\theta_1^*,\theta_3^*)^{-1}\left[\begin{pmatrix} 0 \\ \partial_3 H(z,\theta_3^*)\end{pmatrix}^{\otimes 2}\right]\nu(dz)\\ &= \int 12V(z,\theta_1^*,\theta_3^*)^{-1}\left[\big(\partial_3 H(z,\theta_3^*)\big)^{\otimes 2}\right]\nu(dz)\\ &= \int 12\partial_3 H(z,\theta_3^*)^\star V(z,\theta_1^*,\theta_3^*)^{-1}\partial_3 H(z,\theta_3^*)\nu(dz). \end{aligned}\tag{3.7}$$

The following theorem provides asymptotic normality of $\hat{\theta}_3^0$. The convergence of $\hat{\theta}_3^0$ is much faster than other components of estimators. The proof of the following theorem and the definition of $M_n^{(3)}$ are in Section 6.3.

Theorem 3.3. *Suppose that* [A1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,3,2)$, [A2], [A3] *(iii) and* [A4] *are satisfied. Then*

$$n^{1/2}h^{-1/2}\big(\hat{\theta}_3^0 - \theta_3^*\big) - \Gamma_{33}^{-1}M_n^{(3)} \to^p 0$$

as $n\to\infty$. *In particular,*

$$n^{1/2}h^{-1/2}\big(\hat{\theta}_3^0 - \theta_3^*\big) \to^d N(0,\Gamma_{33}^{-1})$$

as $n\to\infty$.

3.4. About initial estimators

Let

$$\mathbb{H}_n^{(1)}(\theta_1) = -\frac{1}{2}\sum_{j=1}^n \left\{C(Z_{t_{j-1}},\theta_1)^{-1}\left[h^{-1}(\Delta_j X)^{\otimes 2}\right] + \log\det C(Z_{t_{j-1}},\theta_1)\right\}$$

where $\Delta_j X = X_{t_j} - X_{t_{j-1}}$. It should be remarked that the present $\mathbb{H}_n^{(1)}(\theta_1)$ is different from the one given in (4.2) on p. 1436. Under $[A1]$ and $[A2]$ (iii), $\mathbb{H}_n^{(1)}$ is a continuous function on $\overline{\Theta}_1$ a.s.

Given the data $(Z_{t_j})_{j=0,1,\dots,n}$, let us consider the quasi-maximum likelihood estimator (QMLE) $\hat{\theta}_1^0 = \hat{\theta}_{1,n}^0$ for θ_1, that is, $\hat{\theta}_1^0$ is any measurable function of $(Z_{t_j})_{j=0,1,\dots,n}$ satisfying

$$\mathbb{H}_n^{(1)}(\hat{\theta}_1^0) = \max_{\theta_1 \in \overline{\Theta}_1} \mathbb{H}_n^{(1)}(\theta_1) \quad a.s.$$

Routinely, $n^{1/2}$-consistency and asymptotic normality of $\hat{\theta}_1^0$ can be established. We will give a brief for self-containedness and for the later use. Let

$$\Gamma^{(1)}[u_1^{\otimes 2}] = \frac{1}{2} \int_{\mathbb{R}^{\mathsf{d}_Z}} \mathrm{Tr}\big\{C^{-1}(\partial_1 C)[u_1]C^{-1}(\partial_1 C)[u_1](z, \theta_1^*)\big\}\nu(dz) \tag{3.8}$$

for $u_1 \in \mathbb{R}^{\mathsf{p}_1}$. We will see the existence and positivity of $\Gamma^{(1)}$ in the following theorem. We refer the reader to Gloter and Yoshida [22] for a proof.

Theorem 3.4. **(a)** *Suppose that* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (0, 0, 1, 1, 0, 0)$, $[A2]$ *(i), (ii), (iii), and* $[A3]$ *(i) are satisfied. Then* $\hat{\theta}_1^0 \to^p \theta_1^*$ *as* $n \to \infty$.

(b) *Suppose that* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1, 0, 2, 3, 0, 0)$, $[A2]$ *(i), (ii), (iii), and* $[A3]$ *(i) are satisfied. Then* $\Gamma^{(1)}$ *exists and is positive-definite, and*

$$\sqrt{n}\big(\hat{\theta}_1^0 - \theta_1^*\big) - (\Gamma^{(1)})^{-1}\hat{M}_n^{(1)} \to^p 0$$

as $n \to \infty$, *where*

$$\begin{aligned} \hat{M}_n^{(1)} &= \frac{1}{2} n^{-1/2} \sum_{j=1}^n \big(C^{-1}(\partial_1 C)C^{-1}\big)(Z_{t_{j-1}}, \theta_1^*) \\ &\qquad \cdot \big[\big(h^{-1/2}B(Z_{t_{j-1}}, \theta_1^*)\Delta_j w\big)^{\otimes 2} - C(Z_{t_{j-1}}, \theta_1^*)\big]. \end{aligned}$$

Moreover, $M_n^{(1)} \to^d N_{\mathsf{p}_1}(0, \Gamma^{(1)})$ *as* $n \to \infty$. *In particular,*

$$\sqrt{n}\big(\hat{\theta}_1^0 - \theta_1^*\big) \to^d N_{\mathsf{p}_1}\big(0, (\Gamma^{(1)})^{-1}\big)$$

as $n \to \infty$.

Remark 3.5. It is possible to show that the quasi-Bayesian estimator (QBE) also enjoys the same asymptotic properties as the QMLE in Theorem 3.4, if we follows the argument in Yoshida [52]. This means we can use both estimators together with the estimator for θ_2 e.g. given in Section 3.4, to construct a one-step estimator for θ_3 based on the scheme presented in Section 3.1, and consequently we can construct a one-step estimator for $\theta = (\theta_1, \theta_2, \theta_3)$ by the method in Section 4.

We will recall a standard construction of estimator for θ_2. As usual, the scheme is adaptive. Suppose that an estimator $\hat{\theta}_1^0$ based on the data $(Z_{t_j})_{j=0,1,\dots,n}$ satisfies Condition $[A4]$ (i), i.e.,

$$\hat{\theta}_1^0 - \theta_1^* = O_p(n^{-1/2})$$

as $n \to \infty$. Obviously we can apply the estimator constructed above, but any estimator satisfying this condition can be used.

Define the random field $\mathbb{H}_n^{(2)}$ on $\overline{\Theta}_2$ by

$$\mathbb{H}_n^{(2)}(\theta_2) = -\frac{1}{2}\sum_{j=1}^{n} C(Z_{t_{j-1}}, \hat{\theta}_1^0)^{-1}\big[h^{-1}\big(\Delta_j X - hA(Z_{t_{j-1}}, \theta_2)\big)^{\otimes 2}\big]. \tag{3.9}$$

We will denote by $\hat{\theta}_2^0 = \hat{\theta}_{2,n}^0$ any sequence of quasi-maximum likelihood estimator for $\mathbb{H}_n^{(2)}$, that is,

$$\mathbb{H}_n^{(2)}(\hat{\theta}_2^0) = \sup_{\theta_2 \in \overline{\Theta}_2} \mathbb{H}_n^{(2)}(\theta_2).$$

Let $\mathbb{Y}_n^{(2)}(\theta_2) = T^{-1}\big(\mathbb{H}_n^{(2)}(\theta_2) - \mathbb{H}_n^{(2)}(\theta_2^*)\big)$, where $T = nh$. The matrix Γ_{22} is defined by (4.1). Let

$$\hat{M}_n^{(2)} \quad = \quad T^{-1/2}\sum_{j=1}^{n} C(Z_{t_{j-1}}, \theta_1^*)^{-1}\big[B(Z_{t_{j-1}}, \theta_1^*)\Delta_j w, \partial_2 A(Z_{t_{j-1}}, \theta_2^*)\big] \tag{3.10}$$

See Gloter and Yoshida [22] for a proof of the following theorem.

Theorem 3.6. **(a)** *Suppose that Conditions* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,0,0)$, $[A2]$, $[A3]$ *(ii) and* $[A4]$ *(i). Then* $\hat{\theta}_2^0 \to^p \theta_2^*$ *as* $n \to \infty$.

(b) *Suppose that Conditions* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,3,2,1,0,0)$, $[A2]$, $[A3]$ *(ii) and* $[A4]$ *(i). Then*

$$(nh)^{1/2}\big(\hat{\theta}_2^0 - \theta_2^*\big) - \Gamma_{22}^{-1}\hat{M}_n^{(2)} \to^p 0$$

as $n \to \infty$. *In particular,*

$$(nh)^{1/2}\big(\hat{\theta}_2^0 - \theta_2^*\big) \to^d N(0, \Gamma_{22}^{-1})$$

as $n \to \infty$.

4. Adaptive one-step estimator for $(\theta_1, \theta_2, \theta_3)$

In this section, we will consider a one-step estimator for $\theta = (\theta_1, \theta_2, \theta_3)$ given an initial estimators $(\hat{\theta}_1^0, \hat{\theta}_2^0, \hat{\theta}_3^0)$ for $(\theta_1, \theta_2, \theta_3)$ based on $(Z_{t_j})_{j=0,1,\dots,n}$. We will assume the following rate of convergence for each initial estimator.

[A4$^\sharp$] (i) $\hat{\theta}_1^0 - \theta_1^* \;=\; O_p(n^{-1/2})$ as $n \to \infty$

(ii) $\hat{\theta}_2^0 - \theta_2^* \;=\; O_p(n^{-1/2}h^{-1/2})$ as $n \to \infty$

(iii) $\hat{\theta}_3^0 - \theta_3^* \;=\; O_p(n^{-1/2}h^{1/2})$ as $n \to \infty$.

Condition $[A4^\sharp]$ does not assume each initial estimator is attaining the optimal asymptotic variance, nor asymptotically normal. The quasi-maximum likelihood estimator $\hat{\theta}_2^0$ with respect to (3.9) is an option. Another choice of the initial estimator $\hat{\theta}_2^0$ is the simple least squares estimator using the coefficient A though it is less efficient than the quasi-maximum likelihood estimator for the first component of the model. Theorem 4.1 in this section shows the one-step estimator for θ_2 recovers efficiency even if such a less efficient estimator is used as the initial estimator for θ_2.

The initial estimator $\hat{\theta}_3^0$ is not necessarily the one defined in Section 3, though we already know that one satisfies $[A4^\sharp]$ (iii). That is, the initial estimator $\hat{\theta}_3^0$ used in this section is requested to attain the error rate $n^{-1/2}h^{1/2}$ only, not to necessarily achieve the asymptotic variance equal to Γ_{33}^{-1} or less. We know there is an estimator of θ_1 satisfying Condition $[A4^\sharp]$ (i) based on only the first equation of (1.1). It is known that its information cannot be greater than the matrix

$$\frac{1}{2}\int \mathrm{Tr}\big\{\big(\mathrm{C}^{-1}(\partial_1\mathrm{C})\mathrm{C}^{-1}\partial_1\mathrm{C}\big)(\mathrm{z},\theta_1^*)\big\}\nu(\mathrm{dz}).$$

It will be turned out that the amount of information is increased by the one-step estimator.

Let

$$\begin{aligned}\Gamma_{11} &= \frac{1}{2}\int \mathrm{Tr}\big\{\mathrm{S}^{-1}(\partial_1\mathrm{S})\mathrm{S}^{-1}\partial_1\mathrm{S}(\mathrm{z},\theta_1^*,\theta_3^*)\big\}\nu(\mathrm{dz})\\ &= \frac{1}{2}\int \Big[\mathrm{Tr}\big\{\big(\mathrm{C}^{-1}(\partial_1\mathrm{C})\mathrm{C}^{-1}\partial_1\mathrm{C}\big)(\mathrm{z},\theta_1^*)\big\}\\ &\qquad +\mathrm{Tr}\big\{\big(\mathrm{V}^{-1}\mathrm{H_x}(\partial_1\mathrm{C})\mathrm{H_x^\star}\mathrm{V}^{-1}\mathrm{H_x}(\partial_1\mathrm{C})\mathrm{H_x^\star}\big)(\mathrm{z},\theta_1^*,\theta_3^*)\big\}\Big]\nu(\mathrm{dz}).\end{aligned}$$

If H_x is an invertible (square) matrix, then Γ_{11} coincides with

$$\int \mathrm{Tr}\big\{\big(\mathrm{C}^{-1}(\partial_1\mathrm{C})\mathrm{C}^{-1}\partial_1\mathrm{C}\big)(\mathrm{z},\theta_1^*)\big\}\nu(\mathrm{dz}).$$

Otherwise, it is not always true.

Let

$$\begin{aligned}\Gamma_{22} &= \int S(z,\theta_1^*,\theta_3^*)^{-1}\left[\begin{pmatrix}\partial_2 A(z,\theta_2^*)\\ 2^{-1}\partial_2 L_H(z,\theta_1^*,\theta_2^*,\theta_3^*)\end{pmatrix}^{\otimes 2}\right]\nu(dz)\\ &= \int \partial_2 A(z,\theta_2^*)^\star C(z,\theta_1^*)^{-1}\partial_2 A(z,\theta_2^*)\nu(dz).\end{aligned} \tag{4.1}$$

Let $\Gamma^J(\theta^*) = \mathrm{diag}\big[\Gamma_{11},\Gamma_{22},\Gamma_{33}\big]$, where Γ_{33} is defined by (3.7).

We will use the following random fields:

$$\mathbb{H}_n^{(1)}(\theta_1) = -\frac{1}{2}\sum_{j=1}^{n}\Big\{S(Z_{t_{j-1}},\theta_1,\hat{\theta}_3^0)^{-1}\big[\mathcal{D}_j(\theta_1,\hat{\theta}_2^0,\hat{\theta}_3^0)^{\otimes 2}\big] + \log\det S(Z_{t_{j-1}},\theta_1,\hat{\theta}_3^0)\Big\} \tag{4.2}$$

and

$$\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3) = -\frac{1}{2}\sum_{j=1}^{n}\hat{S}(Z_{t_{j-1}},\hat{\theta}_3^0)^{-1}\big[\mathcal{D}_j(\hat{\theta}_1^0,\theta_2,\theta_3)^{\otimes 2}\big]. \tag{4.3}$$

Recall $\hat{S}(z,\theta_3) = S(z,\hat{\theta}_1^0,\theta_3)$. To construct one-step estimators, we consider the functions

$$\mathbb{E}_n(\theta_1) = \theta_1 - \big[\partial_1^2\mathbb{H}_n^{(1)}(\theta_1)\big]^{-1}\partial_1\mathbb{H}_n^{(1)}(\theta_1)$$

and

$$\mathbb{F}_n(\theta_2,\theta_3) = \begin{pmatrix}\theta_2\\ \theta_3\end{pmatrix} - \big[\partial^2_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}\big(\theta_2,\theta_3\big)\big]^{-1}\partial_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}\big(\theta_2,\theta_3\big)$$

when both matrices $\partial_1^2\mathbb{H}_n^{(1)}(\theta_1)$ and $\partial^2_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}\big(\theta_2,\theta_3\big)$ are invertible. Let

$$\mathcal{X}_n^{(1)} = \big\{\omega\in\Omega;\ \partial_1^2\mathbb{H}_n^{(1)}(\hat{\theta}_1^0) \text{ is invertible and } \mathbb{E}_n(\hat{\theta}_1^0)\in\Theta_1\big\}$$

and

$$\mathcal{X}_n^{(2,3)} = \big\{\omega\in\Omega;\ \partial^2_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}\big(\hat{\theta}_2^0,\hat{\theta}_3^0\big) \text{ is invertible and } \mathbb{F}_n(\hat{\theta}_2^0,\hat{\theta}_3^0)\in\Theta_2\times\Theta_3\big\}.$$

Let $\mathcal{X}_n = \mathcal{X}_n^{(1)}\cap\mathcal{X}_n^{(2,3)}$. The event $\mathcal{X}_n$ is a statistic because it is determined by the data $(Z_{t_j})_{j=0,\dots,n}$ only. For $(\theta_1,\theta_2,\theta_3)$, the one-step estimator $(\hat{\theta}_1,\hat{\theta}_2,\hat{\theta}_3)$ with the initial estimator $(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0)$ is defined by

$$\begin{pmatrix}\hat{\theta}_1\\ \hat{\theta}_2\\ \hat{\theta}_3\end{pmatrix} = \begin{cases}\begin{pmatrix}\mathbb{E}_n(\hat{\theta}_1^0)\\ \mathbb{F}_n(\hat{\theta}_2^0,\hat{\theta}_3^0)\end{pmatrix} & \text{on } \mathcal{X}_n\\ \upsilon & \text{on } \mathcal{X}_n^c\end{cases}$$

where υ is an arbitrary value in Θ.

Let

$$\hat{\gamma} = \big(\hat{\theta}_2,\hat{\theta}_3\big)^\star,\quad \hat{\gamma}^0 = \big(\hat{\theta}_2^0,\hat{\theta}_3^0\big)^\star \quad\text{and}\quad \gamma^* = \big(\theta_2^*,\theta_3^*\big)^\star.$$

Let

$$b_n = \begin{pmatrix} n^{-1/2} & 0 & 0\\ 0 & n^{-1/2}h^{-1/2} & 0\\ 0 & 0 & n^{-1/2}h^{1/2}\end{pmatrix}.$$

We obtain a limit theorem for the joint adaptive one-step estimator.

Theorem 4.1. *Suppose that* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,3,2,3,3,2)$*,* $[A2]$*,* $[A3]$ *and* $[A4^\sharp]$ *are satisfied. Then*

$$b_n^{-1}(\hat{\theta} - \theta^*) \to^d N(0, (\Gamma^J(\theta^*))^{-1})$$

as $n \to \infty$*.*

Condition $[A3]$ is used to ensure non-degeneracy of the information matrix. We will give a proof to Theorem 4.1 in Section 7.

5. Basic estimation of the increments

The following sections will be devoted to the proofs.

We have

$$\begin{aligned} h^{-1/2}\Delta_j X &= h^{-1/2}\int_{t_{j-1}}^{t_j} B(Z_t, \theta_1^*)dw_t + h^{-1/2}\int_{t_{j-1}}^{t_j} A(Z_t, \theta_2^*)dt \\ &= h^{-1/2} B(Z_{t_{j-1}}, \theta_1^*)\Delta_j w + r_j^{(5.2)} \end{aligned} \tag{5.1}$$

where

$$\begin{aligned} r_j^{(5.2)} &= h^{-1/2}\int_{t_{j-1}}^{t_j} (B(Z_t, \theta_1^*) - B(Z_{t_{j-1}}, \theta_1^*))dw_t \\ &\quad + h^{-1/2}\int_{t_{j-1}}^{t_j} A(Z_t, \theta_2^*)dt \end{aligned} \tag{5.2}$$

Lemma 5.1. (a) *Under* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (0,0,0,0,0,0)$ *and* $[A2]$ (i)*,*

$$\sup_{s,t\in\mathbb{R}_+,\ |s-t|\le\Delta} \|Z_s - Z_t\|_p = O(\Delta^{1/2}) \quad (\Delta \downarrow 0) \tag{5.3}$$

for every $p > 1$*.*

(b) *Under* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (0,0,1,0,0,0)$ *and* $[A2]$ (i)*,*

$$\sup_n \sup_j \|r_j^{(5.2)}\|_p = O(h^{1/2})$$

for every $p > 1$*.*

Proof. (a) is trivial. For (b), the first term on the right-hand side of (5.2) can be estimated by the Burkholder-Davis-Gundy inequality, Taylor's formula for $B(Z_t, \theta_1^*) - B(Z_{t_{j-1}}, \theta_1^*)$ and by (5.3). □

We have

$$h^{-1/2}\Delta_j X = h^{-1/2}\int_{t_{j-1}}^{t_j} B(Z_t, \theta_1^*)dw_t + h^{1/2} A(Z_{t_{j-1}}, \theta_2^*) + r_j^{(5.4)}$$

where

$$r_j^{(5.4)} = h^{-1/2}\int_{t_{j-1}}^{t_j} \big(A(Z_t, \theta_2^*) - A(Z_{t_{j-1}}, \theta_2^*)\big)dt. \tag{5.4}$$

Then, thanks to (5.3), we obtain the following estimate.

Lemma 5.2. *$r_j^{(5.4)} = O_{L^{\infty-}}(h)$ uniformly, i.e.,*

$$\sup_n \sup_j \left\| r_j^{(5.4)} \right\|_p = O(h) \tag{5.5}$$

for every $p > 1$ if [A1] for $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,0,0,0,0,0)$ and [A2] (i) hold.

Write

$$\zeta_j = \sqrt{3} \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} dw_s dt.$$

Then $E\left[\zeta_j^{\otimes 2}\right] = h^3 I_{\mathsf{r}}$ for the r-dimensional identity matrix I_{r}.

The function G_n is defined in (3.1). Under sufficient smoothness of the coefficients, we have

$$\begin{aligned}
& \Delta_j Y - hG_n(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3) \\
= \; & \Delta_j Y - hH(Z_{t_{j-1}}, \theta_3) - \frac{h^2}{2} L_H\big(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3\big) \\
= \; & hH(Z_{t_{j-1}}, \theta_3^*) - hH(Z_{t_{j-1}}, \theta_3) \\
& + H_x(Z_{t_{j-1}}, \theta_3^*) B(Z_{t_{j-1}}, \theta_1^*) \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} dw_s dt \\
& + \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \big\{ H_x(Z_s, \theta_3^*) B(Z_s, \theta_1^*) - H_x(Z_{t_{j-1}}, \theta_3^*) B(Z_{t_{j-1}}, \theta_1^*) \big\} dw_s dt \\
& + \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \big(L_H(Z_s, \theta_1^*, \theta_2^*, \theta_3^*) - L_H(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3) \big) ds dt \\
= \; & \big\{ hH(Z_{t_{j-1}}, \theta_3^*) - hH(Z_{t_{j-1}}, \theta_3) \big\} + \kappa(Z_{t_{j-1}}, \theta_1^*, \theta_3^*) \zeta_j \\
& + \rho_j(\theta_1, \theta_2, \theta_3)
\end{aligned} \tag{5.6}$$

where $\kappa(Z_{t_{j-1}}, \theta_1^*, \theta_3^*)$ is given in (3.4).

$$\begin{aligned}
\rho_j(\theta_1, \theta_2, \theta_3) \; = \; & \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \big\{ H_x(Z_s, \theta_3^*) B(Z_s, \theta_1^*) \\
& \qquad - H_x(Z_{t_{j-1}}, \theta_3^*) B(Z_{t_{j-1}}, \theta_1^*) \big\} dw_s dt \\
& + \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \big(L_H(Z_s, \theta_1^*, \theta_2^*, \theta_3^*) - L_H(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3) \big) ds dt.
\end{aligned} \tag{5.7}$$

Lemma 5.3. *Suppose that [A1] with $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,0,1,0,3,0)$ and [A2] (i) are satisfied. Then*

(a) $\sup_n \sup_j \left\| \rho_j(\theta_1^*, \theta_2^*, \theta_3^*) \right\|_p = O(h^2)$ *for every $p > 1$.*

(b) $\sup_n \sup_j \left\| \mathcal{D}_j(\theta_1^*, \theta_2^*, \theta_3^*) \right\|_p < \infty$ *for every $p > 1$.*

Proof. It is possible to show (a) by (5.7) and using the estimate (5.3) with the help of Taylor's formula. Additionally to the representation (5.6), by using (5.1) and (5.2), we obtain (b). □

We denote by $(B_xB)(z,\theta_2)$ the tensor defined by $(B_xB)(z,\theta_2)[u_1 \otimes u_2] = B_x(z,\theta_2)[u_2, B(z,\theta_2)[u_1]]$ for $u_1, u_2 \in \mathbb{R}^{\mathsf{r}}$. Moreover, we write $dw_s dw_t$ for $dw_s \otimes dw_t$, and

$$(B_xB)(Z_{t_{j-1}},\theta_2^*)\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t} dw_s dw_t$$

for

$$(B_xB)(Z_{t_{j-1}},\theta_2^*)\Big[\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t} dw_s dw_t\Big].$$

We will apply this rule in similar situations. Let

$$\begin{aligned} L_B(z,\theta_1,\theta_2,\theta_3) \quad = \quad & B_x(z,\theta_1)[A(z,\theta_2)] + \frac{1}{2}B_{xx}(z,\theta_1)[C(z,\theta_1)] \\ & + B_y(z,\theta_3)[H(z,\theta_3)]. \end{aligned} \tag{5.8}$$

Lemma 5.4. *Suppose that* $[A1]$ *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,0,0,0)$ *and* $[A2]$ (i) *are satisfied. Then*

$$h^{-1/2}\big(\Delta_j X - hA(Z_{t_{j-1}},\theta_2)\big) = \xi_j^{(5.10)} + \xi_j^{(5.11)} + r_j^{(5.12)}(\theta_2) \tag{5.9}$$

where

$$\xi_j^{(5.10)} = h^{-1/2}B(Z_{t_{j-1}},\theta_1^*)\Delta_j w, \tag{5.10}$$

$$\xi_j^{(5.11)} = h^{-1/2}(B_xB)(Z_{t_{j-1}},\theta_1^*)\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t} dw_s dw_t, \tag{5.11}$$

and

$$\begin{aligned} r_j^{(5.12)}(\theta_2) \quad = \quad & h^{-1/2}\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t}((B_xB)(Z_s,\theta_1^*) - (B_xB)(Z_{t_{j-1}},\theta_1^*))dw_s dw_t \\ & + h^{-1/2}\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t} L_B(Z_s,\theta_1^*,\theta_2^*,\theta_3^*)ds dw_t \\ & + h^{-1/2}\int_{t_{j-1}}^{t_j}\big(A(Z_t,\theta_2^*) - A(Z_{t_{j-1}},\theta_2)\big)dt. \end{aligned} \tag{5.12}$$

Moreover,

$$\sup_n \sup_j \|r_j^{(5.12)}(\theta_2^*)\|_p = O(h) \tag{5.13}$$

for every $p > 1$, and

$$\left|r_j^{(5.12)}(\theta_2)\right| \le r_{n,j}^{(5.15)}\left\{h^{1/2}\left|\theta_2 - \theta_2^*\right| + h\right\} \tag{5.14}$$

with some random variables $r_{n,j}^{(5.15)}$ satisfying

$$\sup_n \sup_j \left\| r_{n,j}^{(5.15)} \right\|_p < \infty \tag{5.15}$$

for every $p > 1$.

Proof. The decomposition (5.9) is obtained by Itô's formula. The estimate (5.13) is verified by (5.3) since $\partial_z(B_x B)$ and $\partial_z A$ are bound by a polynomial in z uniformly in θ. The estimate (5.14) uses $\partial_2 A$ for θ_2 near θ_2^* as well as $\partial_z A$ evaluated at θ_2^*:

$$\left|r_j^{(5.12)}(\theta_2)\right| 1_{\{|\theta_2-\theta_2^*|<r\}} \le r_{n,j}^{(5.15)}\left\{h^{1/2}\left|\theta_2 - \theta_2^*\right| + h\right\} 1_{\{|\theta_2-\theta_2^*|<r\}}$$

with some positive constant r and some random variables $r_{n,j}^{(5.15)}$ satisfying (5.15). The small number r was taken to ensure convexity of the vicinity of θ_2^*. For θ_2 such that $|\theta_2 - \theta_2^*| \ge r$, the estimate (5.14) is valid by enlarging $r_{n,j}^{(5.15)}$ if necessary. □

Lemma 5.5. **(a)** *Suppose that* [A1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1, 1, 2, 1, 3, 0)$ *and* [A2] (*i*) *are satisfied. Then*

$$\begin{aligned} \Delta_j Y - hG_n(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3^*) &= \xi_j^{(5.17)} + \xi_j^{(5.18)} + h^{3/2} r_j^{(5.19)}(\theta_1, \theta_2) \\ &\quad + h^{3/2} r_j^{(5.20)}(\theta_1, \theta_2) \end{aligned} \tag{5.16}$$

where

$$\xi_j^{(5.17)} = \kappa(Z_{t_{j-1}}, \theta_1^*, \theta_3^*)\zeta_j, \tag{5.17}$$

$$\xi_j^{(5.18)} = ((H_x B)_x B)(Z_{t_{j-1}}, \theta_1^*, \theta_3^*) \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \int_{t_{j-1}}^{s} dw_r dw_s dt, \tag{5.18}$$

$$\begin{aligned} r_j^{(5.19)}(\theta_1, \theta_2) &= h^{-3/2} \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \int_{t_{j-1}}^{s} \big\{((H_x B)_x B)(Z_r, \theta_1^*, \theta_3^*) \\ &\quad - ((H_x B)_x B)(Z_{t_{j-1}}, \theta_1^*, \theta_3^*)\big\} dw_r dw_s dt \\ &\quad + h^{-3/2} \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \int_{t_{j-1}}^{s} L_{H_x B}(Z_r, \theta_1^*, \theta_2^*, \theta_3^*) dr dw_s dt \\ &\quad + h^{-3/2} \int_{t_{j-1}}^{t_j} \int_{t_{j-1}}^{t} \big(L_H(Z_s, \theta_1, \theta_2, \theta_3^*) \\ &\qquad - L_H(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3^*)\big) ds dt \end{aligned} \tag{5.19}$$

with

$$\begin{aligned} L_{H_xB}(z,\theta_1,\theta_2,\theta_3) &= (H_xB)_x(z,\theta_1,\theta_3)[A(z,\theta_2)] \\ &\quad +\frac{1}{2}(H_xB)_{xx}(z,\theta_1,\theta_3)[C(z,\theta_1)] \\ &\quad +(H_xB)_y(z,\theta_1,\theta_3)[H(z,\theta_3)], \end{aligned}$$

and

$$\begin{aligned} r_j^{(5.20)}(\theta_1,\theta_2) &= h^{-3/2}\int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t}\big(L_H(Z_s,\theta_1^*,\theta_2^*,\theta_3^*) \\ &\qquad -L_H(Z_s,\theta_1,\theta_2,\theta_3^*)\big)dsdt. \end{aligned} \tag{5.20}$$

Moreover,

$$\sup_n\sup_j\Big\|\sup_{(\theta_1,\theta_2)\in\overline{\Theta}_1\times\overline{\Theta}_2}\big|r_j^{(5.19)}(\theta_1,\theta_2)\big|\Big\|_p = O(h) \tag{5.21}$$

for every $p>1$, and

$$\big|r_j^{(5.20)}(\theta_1,\theta_2)\big| \leq h^{1/2}r_{n,j}^{(5.23)}\big\{\big|\theta_1-\theta_1^*\big|+\big|\theta_2-\theta_2^*\big|\big\} \tag{5.22}$$

for all $(\theta_1,\theta_2)\in\overline{\Theta}_1\times\overline{\Theta}_2$ with some random variables $r_{n,j}^{(5.23)}$ satisfying

$$\sup_n\sup_j\big\|r_{n,j}^{(5.23)}\big\|_p<\infty \tag{5.23}$$

for every $p>1$.

(b) *Suppose that $[A1]$ with $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,2,0)$ and $[A2]$ (i) are satisfied. Then there exist random variables $r_{n,j}^{(5.24)}$ and a number ρ such that*

$$\sup_{\theta_3\in\overline{\Theta}_3}\big|\mathcal{D}_j(\theta_1,\theta_2,\theta_3)-\mathcal{D}_j(\theta_1^*,\theta_2^*,\theta_3)\big|\leq h^{1/2}r_{n,j}^{(5.24)}\big\{\big|\theta_1-\theta_1^*\big|+\big|\theta_2-\theta_2^*\big|\big\}$$

for all $(\theta_1,\theta_2)\in B((\theta_1^*,\theta_2^*),\rho)$ and that

$$\sup_n\sup_j\big\|r_{n,j}^{(5.24)}\big\|_p<\infty \tag{5.24}$$

for every $p>1$.

Proof. By (5.6), we have

$$\Delta_jY-hG_n(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3^*)=\xi_j^{(5.17)}+\rho_j(\theta_1,\theta_2,\theta_3^*) \tag{5.25}$$

and

$$\rho_j(\theta_1,\theta_2,\theta_3^*) = \int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t}\big\{H_x(Z_s,\theta_3^*)B(Z_s,\theta_1^*) - H_x(Z_{t_{j-1}},\theta_3^*)B(Z_{t_{j-1}},\theta_1^*)\big\}dw_s dt + \int_{t_{j-1}}^{t_j}\int_{t_{j-1}}^{t}\big(L_H(Z_s,\theta_1^*,\theta_2^*,\theta_3^*) - L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3^*)\big)ds dt.$$

Then the decomposition (5.16) is obvious. The first and third terms on the right-hand side of (5.19) can be estimated with Taylor's formula and (5.3), and the second term is easy to estimate. Thus, we obtain (5.21). Since $\partial_{(\theta_1,\theta_2)}L_H(z,\theta_1,\theta_2,\theta_3^*)$ is bounded by a polynomial in z uniformly in (θ_1,θ_2), there exist random variables $r_{n,j}^{(5.23)}$ that satisfy (5.22) and (5.23). [First show (5.22) on the set $\{|(\theta_1,\theta_2)-(\theta_1^*,\theta_2^*)|<r\}$, next see this estimate is valid on $(\overline{\Theta}_1\times\overline{\Theta}_2)\setminus\{|(\theta_1,\theta_2)-(\theta_1^*,\theta_2^*)|<r\}$ by redefining $r_{n,j}^{(5.23)}$ if necessary.] We obtained (a). The assertion (b) is easy to verify with (5.6), (5.7) and Lemma 5.4. □

Lemma 5.6. *Suppose that* [A1] *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(0,0,0,0,2,1)$ *and* [A2] (*i*) *are satisfied. Then*

$$\sup_{(\theta_1,\theta_2)\in\overline{\Theta}_1\times\overline{\Theta}_2}\left|\mathcal{D}_j(\theta_1,\theta_2,\theta_3)-\mathcal{D}_j(\theta_1,\theta_2,\theta_3')\right| \le h^{-1/2}r_{n,j}^{(5.26)}\left|\theta_3-\theta_3'\right| \qquad (\theta_3,\theta_3'\in\overline{\Theta}_3)$$

for some random variables $r_{n,j}^{(5.26)}$ *such that*

$$\sup_n\sup_j\left\|r_{n,j}^{(5.26)}\right\|_p<\infty \tag{5.26}$$

for every $p>1$.

Proof.

$$\mathcal{D}_j(\theta_1,\theta_2,\theta_3)-\mathcal{D}_j(\theta_1,\theta_2,\theta_3') = \left(\begin{array}{c} 0 \\ \left\{\begin{array}{c} h^{-1/2}\big(H(Z_{t_{j-1}},\theta_3')-H(Z_{t_{j-1}},\theta_3)\big) \\ +\frac{h^{1/2}}{2}\big(L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3')-L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\big)\end{array}\right\}\end{array}\right)$$

Therefore the lemma is obvious. Apply the Taylor formula for the argument θ_3 if θ_3 and θ_3' are close, otherwise and if necessary, redifine $r_{n,j}^{(5.26)}$. □

6. Proof of Theorems 3.2 and 3.3

6.1. Proof of Theorem 3.2

Lemma 6.1. *Suppose that* [*A*1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (0,0,0,1,1,1)$ *and* [*A*2] *(i), (iii) and (iv) are fulfilled. Then*

$$\sup_{t\in\mathbb{R}_+}\bigg\|\sup_{(\theta_1,\theta_3)\in\overline{\Theta}_1\times\overline{\Theta}_3}\Big\{\big|S(Z_t,\theta_1,\theta_3)\big| + \det S(Z_t,\theta_1,\theta_3)^{-1} + \big|S(Z_t,\theta_1,\theta_3)^{-1}\big|\Big\}\bigg\|_p < \infty$$

for every $p > 1$

Proof. By [*A*2] (iii) and (iv), $\det S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}$ as well as $S(Z_{t_{j-1}},\theta_1,\theta_3)$ is continuous on $\overline{\Theta}_1\times\overline{\Theta}_3$ a.s., and continuously differentiable on $\Theta_1\times\Theta_3$. Moreover we see

$$\sup_{t\in\mathbb{R}_+}\sum_{i=0,1}\sup_{(\theta_1,\theta_3)\in\Theta_1\times\Theta_3}\big\|\partial^i_{(\theta_1,\theta_3)}\big(\det S(Z_t,\theta_1,\theta_3)^{-1}\big)\big\|_p < \infty$$

for every $p > 1$ from (3.3). This implies that

$$\sup_{t\in\mathbb{R}_+}\bigg\|\sup_{(\theta_1,\theta_3)\in\Theta_1\times\Theta_3}\big(\det S(Z_t,\theta_1,\theta_3)^{-1}\big)\bigg\|_p < \infty$$

for every $p > 1$ by Sobolev's inequality. The inequality

$$\sup_{t\in\mathbb{R}_+}\bigg\|\sup_{(\theta_1,\theta_3)\in\Theta_1\times\Theta_3}\big|S(Z_t,\theta_1,\theta_3)\big|\bigg\|_p < \infty$$

for every $p > 1$ is rather easy to show. □

Proof of Theorem 3.2. We have

$$\mathbb{Y}^{(3)}_n(\theta_3) = -\frac{1}{2n}\sum_{j=1}^n \hat{S}(Z_{t_{j-1}},\theta_3)^{-1}\big[\big(h^{1/2}\delta_j(\hat{\theta}^0_1,\hat{\theta}^0_2,\theta_3)\big)^{\otimes 2}\big] + n^{-1}hR^{(6.1)}_n(\theta_3),$$

where

$$\delta_j(\theta_1,\theta_2,\theta_3) = -\mathcal{D}_j(\theta_1,\theta_2,\theta_3) + \mathcal{D}_j(\theta_1,\theta_2,\theta^*_3)$$

and

$$R^{(6.1)}_n(\theta_3) = h^{-1/2}\sum_{j=1}^n \hat{S}(Z_{t_{j-1}},\theta_3)^{-1}\big[h^{1/2}\delta_j(\hat{\theta}^0_1,\hat{\theta}^0_2,\theta_3),\mathcal{D}_j(\hat{\theta}^0_1,\hat{\theta}^0_2,\theta^*_3)\big] -\frac{1}{2}\sum_{j=1}^n\big(\hat{S}(Z_{t_{j-1}},\theta_3)^{-1} - \hat{S}(Z_{t_{j-1}},\theta^*_3)^{-1}\big)\big[\mathcal{D}_j(\hat{\theta}^0_1,\hat{\theta}^0_2,\theta^*_3)^{\otimes 2}\big]$$

$$-\frac{1}{2}\sum_{j=1}^{n}\log\frac{\det\hat{S}(Z_{t_{j-1}},\theta_3)}{\det\hat{S}(Z_{t_{j-1}},\theta_3^*)} \tag{6.1}$$

By Lemma 5.3 (b), Lemma 5.5 (b), Lemma 5.6 and Lemma 6.1, we obtain

$$n^{-1}h\sup_{\theta_3\in\Theta_3}\left|R_n^{(6.1)}(\theta_3)\right| = O_p(h^{1/2})+O_p(h) = O_p(h^{1/2}).$$

By definition,

$$\begin{aligned} & h^{1/2}\delta_j(\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3) \\ = & \left(\begin{array}{c} 0 \\ \left\{\begin{array}{c} H(Z_{t_{j-1}},\theta_3)-H(Z_{t_{j-1}},\theta_3^*) \\ +\frac{h}{2}\big(L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)-L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3^*)\big) \end{array}\right\} \end{array}\right). \end{aligned}$$

Since the functions $A(z,\theta_2)$, $H(z,\theta_3)$ and $L_H(z,\theta_1,\theta_2,\theta_3)$ are dominated by a polynomial in z uniformly in θ, by using the above formula, it is easy to show

$$\sup_{\theta_3\in\overline{\Theta}_3}\left|\mathbb{Y}_n^{(3)}(\theta_3)-\mathbb{Y}_n^{(6.3)}(\hat{\theta}_1^0,\theta_3)\right| = O_p(h^{1/2}) \tag{6.2}$$

for

$$\begin{aligned} & \mathbb{Y}_n^{(6.3)}(\theta_1,\theta_3) \\ = & -\frac{1}{2n}\sum_{j=1}^{n}S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\left(\begin{array}{c} 0 \\ H(Z_{t_{j-1}}\,\theta_3)-H(Z_{t_{j-1}}\,\theta_3^*) \end{array}\right)^{\otimes 2}\right]. \end{aligned} \tag{6.3}$$

The derivative $\partial_1 S_x(z,\theta_1,\theta_3)$ is dominated by a polynomial in z uniformly in θ. Therefore

$$\sup_{\theta_3\in\overline{\Theta}_3}\left|\mathbb{Y}_n^{(6.3)}(\hat{\theta}_1^0,\theta_3)-\mathbb{Y}_n^{(6.3)}(\theta_1^*,\theta_3)\right| \to^p 0. \tag{6.4}$$

Finally, the estimate (5.3) gives

$$\begin{aligned} \sup_{\theta_3\in\overline{\Theta}_3}\Bigg| & \mathbb{Y}_n^{(6.3)}(\theta_1^*,\theta_3) \\ & +\frac{1}{2nh}\int_0^{nh}S(Z_t,\theta_1^*,\theta_3)^{-1}\left[\left(\begin{array}{c} 0 \\ H(Z_t,\theta_3)-H(Z_t,\theta_3^*) \end{array}\right)^{\otimes 2}\right]dt\Bigg| \\ \to^p 0. & \end{aligned} \tag{6.5}$$

Now (3.6) follows from (6.2), (6.4), (6.5) and [A2] (ii) since ${\partial_3}^i H(z,\theta_1,\theta_3)$ $(i=0,1)$ are dominated by a polynomial in z uniformly in θ_3. Then the convergence $\hat{\theta}_3^0 \to^p \theta_3$ as $n\to\infty$ is obvious under Condition [A3] (iii). □

6.2. Random fields

Let

$$\widetilde{\mathcal{D}}_j(\theta_1',\theta_2',\theta_3') = \mathcal{D}_j(\theta_1',\theta_2',\theta_3') + \widetilde{\mathcal{D}}_j(\theta_1^*,\theta_2^*,\theta_3^*) - \mathcal{D}_j(\theta_1^*,\theta_2^*,\theta_3^*)$$

where

$$\widetilde{\mathcal{D}}_j(\theta_1^*,\theta_2^*,\theta_3^*) = \begin{pmatrix} \xi_j^{(5.10)} + \xi_j^{(5.11)} \\ h^{-3/2}\big(\xi_j^{(5.17)} + \xi_j^{(5.18)}\big) \end{pmatrix}.$$

To solve the problem, we need to exploit stochastic orthogonality between random fields. Though this technique is standard, to carry out it as visibly as possible, we necessarily introduce various random fields below. These symbols are useful to clarify which parameters are replaced in the formula and which order of error is caused, also to make big formulas compact and to avoid repetition of them. The following random fields depend on n.

$$\begin{aligned} &\Psi_{3,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3') \\ = \;& \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\mathcal{D}_j(\theta_1',\theta_2',\theta_3'),\ \begin{pmatrix} 0 \\ \partial_3 H(Z_{t_{j-1}},\theta_3) \end{pmatrix}\right], \end{aligned}$$

$$\begin{aligned} &\widetilde{\Psi}_{3,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3') \\ = \;& \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\widetilde{\mathcal{D}}_j(\theta_1',\theta_2',\theta_3'),\ \begin{pmatrix} 0 \\ \partial_3 H(Z_{t_{j-1}},\theta_3) \end{pmatrix}\right], \end{aligned}$$

$$\begin{aligned} &\Psi_{3,2}(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3') \\ = \;& \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\mathcal{D}_j(\theta_1',\theta_2',\theta_3'),\ \begin{pmatrix} 0 \\ 2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3) \end{pmatrix}\right], \end{aligned}$$

$$\begin{aligned} &\Psi_{3,3}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3') \\ = \;& \frac{1}{2}\sum_{j=1}^n \big(S^{-1}(\partial_3 S)S^{-1}\big)(Z_{t_{j-1}},\theta_1,\theta_3)\big[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')^{\otimes 2} - S(Z_{t_{j-1}},\theta_1,\theta_3)\big], \end{aligned}$$

$$\begin{aligned} &\Psi_{33,1}(\theta_1,\theta_2,\theta_3) \\ = \;& -\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\begin{pmatrix} 0 \\ \left\{\begin{array}{c} \partial_3 H(Z_{t_{j-1}},\theta_3) \\ +2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3) \end{array}\right\} \end{pmatrix}^{\otimes 2}\right], \end{aligned}$$

$$
\begin{aligned}
&\Psi_{33,2}(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\;& \sum_{j=1}^{n} S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad \otimes \begin{pmatrix} 0 \\ \partial_3^2 H(Z_{t_{j-1}},\theta_3)+2^{-1}h\partial_3^2 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\end{pmatrix}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
&\Psi_{33,3}(\theta_1,\theta_3)\\
=\;& -\frac{1}{2}\sum_{j=1}^{n}\big\{\big(S^{-1}(\partial_3 S)S^{-1}\big)(Z_{t_{j-1}},\theta_1,\theta_3)\big[\partial_3 S(Z_{t_{j-1}},\theta_1,\theta_3)\big]\big\},
\end{aligned}
$$

$$
\begin{aligned}
&\Psi_{33,4}(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\;& -2\sum_{j=1}^{n} S^{-1}(\partial_3 S)S^{-1}(Z_{t_{j-1}},\theta_1,\theta_3)\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad \otimes \begin{pmatrix} 0 \\ \partial_3 H(Z_{t_{j-1}},\theta_3)+2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\end{pmatrix}\Bigg],
\end{aligned}
$$

$$
\begin{aligned}
&\Psi_{33,5}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\;& \frac{1}{2}\sum_{j=1}^{n}\partial_3\big\{\big(S^{-1}(\partial_3 S)S^{-1}\big)(Z_{t_{j-1}},\theta_1,\theta_3)\big\}\\
&\qquad \cdot\Big[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')^{\otimes 2}-S(Z_{t_{j-1}},\theta_1,\theta_3)\Big].
\end{aligned}
$$

6.3. Proof of Theorem 3.3

Let

$$
\begin{aligned}
M_n^{(3)} \;=\;& n^{-1/2}\sum_{j=1}^{n} S(Z_{t_{j-1}},\theta_1^*,\theta_3^*)^{-1}\Bigg[\begin{pmatrix} h^{-1/2}B(Z_{t_{j-1}},\theta_2^*)\Delta_j w \\ h^{-3/2}\kappa(Z_{t_{j-1}},\theta_1^*,\theta_3^*)\zeta_j\end{pmatrix}\\
&\qquad \otimes \begin{pmatrix} 0 \\ \partial_3 H(Z_{t_{j-1}},\theta_3^*)\end{pmatrix}\Bigg].
\end{aligned}
$$

Lemma 6.2. *Suppose that* [A1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,3,1)$, [A2] *and* [A4] *are satisfied. Then*

$$
n^{-1/2}h^{1/2}\,\partial_3\mathbb{H}_n^{(3)}(\theta_3^*) - M_n^{(3)} = o_p(1)
$$

as $n\to\infty$.

Proof. From (3.5) and (3.2), we have

$$n^{-1/2}h^{1/2}\,\partial_3\mathbb{H}_n^{(3)}(\theta_3^*) \quad = \quad R_n^{(6.7)}(\hat\theta_1^0,\hat\theta_2^0)+R_n^{(6.8)}(\hat\theta_1^0,\hat\theta_2^0)+R_n^{(6.9)}(\hat\theta_1^0,\hat\theta_2^0) \tag{6.6}$$

where

$$R_n^{(6.7)}(\hat\theta_1^0,\hat\theta_2^0)=n^{-1/2}\Psi_{3,1}(\hat\theta_1^0,\theta_3^*,\hat\theta_1^0,\hat\theta_2^0,\theta_3^*), \tag{6.7}$$

$$R_n^{(6.8)}(\hat\theta_1^0,\hat\theta_2^0) \quad = \quad n^{-1/2}\Psi_{3,2}(\hat\theta_1^0,\hat\theta_2^0,\theta_3^*,\hat\theta_1^0,\hat\theta_2^0,\theta_3^*) \tag{6.8}$$

and

$$R_n^{(6.9)}(\hat\theta_1^0,\hat\theta_2^0)=n^{-1/2}h^{1/2}\Psi_{3,3}(\hat\theta_1^0,\theta_3^*,\hat\theta_1^0,\hat\theta_2^0,\theta_3^*). \tag{6.9}$$

We have

$$\begin{aligned}&\mathcal{D}_j(\hat\theta_1^0,\hat\theta_2^0,\theta_3^*)-\mathcal{D}_j(\hat\theta_1^0,\theta_2^*,\theta_3^*)\\ =\quad& -h^{1/2}\left(\begin{array}{c}\big(A(Z_{t_{j-1}},\hat\theta_2^0)-A(Z_{t_{j-1}},\theta_2^*)\big)\\ 2^{-1}H_x(Z_{t_{j-1}},\theta_3^*)\big[A(Z_{t_{j-1}},\hat\theta_2^0)-A(Z_{t_{j-1}},\theta_2^*)\big]\end{array}\right),\end{aligned}$$

and so only by algebraic computation we obtain

$$\begin{aligned}&\hat S(Z_{t_{j-1}},\theta_3^*)^{-1}\left[\mathcal{D}_j(\hat\theta_1^0,\hat\theta_2^0,\theta_3^*)-\mathcal{D}_j(\hat\theta_1^0,\theta_2^*,\theta_3^*),\ \left(\begin{array}{c}0\\ \partial_3H(Z_{t_{j-1}},\theta_3^*)\end{array}\right)\right]\\ =\quad& 0.\end{aligned} \tag{6.10}$$

Applying Lemma 5.5 (b) under $[A4]$, and next using the results in Lemmas 5.4 and 5.5, we see

$$\begin{aligned}R_n^{(6.7)}(\hat\theta_1^0,\hat\theta_2^0) \quad &= \quad n^{-1/2}\Psi_{3,1}(\hat\theta_1^0,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)+O_p(h^{1/2})\\ &= \quad n^{-1/2}\tilde\Psi_{3,1}(\hat\theta_1^0,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)+o_p(1)\end{aligned} \tag{6.11}$$

since $(nh^2)^{1/2}=o(1)$. Consider the random field

$$\Phi_n^{(6.12)}(u_1) \quad = \quad n^{-1/2}\big\{\widetilde\Psi_{3,1}(\theta_1^*+r_nu_1,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)-\widetilde\Psi_{3,1}(\theta_1^*,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)\big\} \tag{6.12}$$

on $\{u_1\in\mathbb{R}^{\mathsf{p}_1};\ |u_1|<1\}$ for any sequence of positive numbers $r_n\to 0$. Sobolev's inequality gives

$$\sup_{u_1:|u_1|<1}|\Phi_n^{(6.12)}(u_1)|=o_p(1)$$

with the help of orthogonality. In particular,

$$R_n^{(6.7)}(\hat{\theta}_1^0, \hat{\theta}_2^0) \quad = \quad n^{-1/2}\widetilde{\Psi}_{3,1}(\theta_1^*, \theta_3^*, \theta_1^*, \theta_2^*, \theta_3^*) + o_p(1). \tag{6.13}$$

This implies

$$R_n^{(6.7)}(\hat{\theta}_1^0, \hat{\theta}_2^0) = M_n^{(3)} + o_p(1).$$

Simpler is that $R_n^{(6.8)}(\hat{\theta}_1^0, \hat{\theta}_2^0) = O_p(n^{1/2}h)$. Similarly,

$$\begin{aligned} R_n^{(6.9)}(\hat{\theta}_1^0, \hat{\theta}_2^0) &= n^{-1/2}h^{1/2}\Psi_{3,3}(\theta_1^*, \theta_3^*, \theta_1^*, \theta_2^*, \theta_3^*) + O_p(h^{1/2}) \\ &= O_p(h^{1/2}). \end{aligned}$$

Thus, we obtained the result. □

In what follows, we quite often use the estimates in Lemma 6.1 without mentioning it explicitly.

Lemma 6.3. *Suppose that [A1] with $(i_A, j_A, i_B, j_B, i_H, j_H) = (1, 1, 2, 1, 3, 2)$, [A2] and [A4] are satisfied. Then*

$$\sup_{\theta_3 \in B_n} \left| n^{-1}h\, \partial_3^2 \mathbb{H}_n^{(3)}(\theta_3) + \Gamma_{33} \right| \to^p 0$$

for any sequence of balls B_n in $\mathbb{R}^{\mathsf{p}_3}$ shrinking to θ_3^*.

Proof. We have

$$\begin{aligned} n^{-1}h\, \partial_3^2 \mathbb{H}_n^{(3)}(\theta_3) \quad = \quad & n^{-1}\Psi_{33,1}(\hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3) \\ & + n^{-1}h^{1/2}\Psi_{33,2}(\hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3, \hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3) \\ & + n^{-1}h\Psi_{33,3}(\hat{\theta}_1^0, \theta_3) \\ & + n^{-1}h^{1/2}\Psi_{33,4}(\hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3, \hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3) \\ & + n^{-1}h\Psi_{33,5}(\hat{\theta}_1^0, \theta_3, \hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3). \end{aligned}$$

For $\mathcal{D}_j(\hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3)$ in the above expression, we use Lemma 5.5 (b) to replace $\hat{\theta}_i^0$ by θ_i^* for $i = 1, 2$, and Lemma 5.6 to replace $\theta_3 \in B_n$ by θ_3^* with an error uniform in $\theta_3 \in B_n$. Next we use Lemma 5.3 (b). Then

$$\begin{aligned} n^{-1}h\, \partial_3^2 \mathbb{H}_n^{(3)}(\theta_3) \quad = \quad & -n^{-1}\sum_{j=1}^n \hat{S}(Z_{t_{j-1}}, \theta_3)^{-1}\left[\begin{pmatrix} 0 \\ \partial_3 H(Z_{t_{j-1}}, \theta_3) \end{pmatrix}^{\otimes 2}\right] \\ & + r_n^{(6.14)}(\theta_3) \end{aligned}$$

where

$$\sup_{\theta_3 \in \overline{\Theta}_3} \left| r_n^{(6.14)}(\theta_3) \right| = o_p(1). \tag{6.14}$$

Now we obtain the result by using [A2] and estimating the functions $\partial_3 S$ and $\partial_3^2 H$ uniformly in (θ_1, θ_3). □

Now Theorem 3.3 follows from Lemmas 6.2 and 6.3.

7. Proof of Theorem 4.1

Let us prepare the following random fields.

$$\begin{aligned}
&\Psi_2(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\qquad\otimes\begin{pmatrix}\partial_2 A(Z_{t_{j-1}},\theta_2)\\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\end{pmatrix}\Bigg]\\
=\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\qquad\otimes\begin{pmatrix}\partial_2 A(Z_{t_{j-1}},\theta_2)\\ 2^{-1}H_x(z,\theta_3)[\partial_2 A(Z_{t_{j-1}},\theta_2)]\end{pmatrix}\Bigg],
\end{aligned}$$

$$\begin{aligned}
&\widetilde{\Psi}_2(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\widetilde{\mathcal{D}}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\qquad\otimes\begin{pmatrix}\partial_2 A(Z_{t_{j-1}},\theta_2)\\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\end{pmatrix}\Bigg]\\
=\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\widetilde{\mathcal{D}}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\qquad\otimes\begin{pmatrix}\partial_2 A(Z_{t_{j-1}},\theta_2)\\ 2^{-1}H_x(z,\theta_3)[\partial_2 A(Z_{t_{j-1}},\theta_2)]\end{pmatrix}\Bigg],
\end{aligned}$$

$$\begin{aligned}
&\Psi_3(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\otimes\begin{pmatrix}0\\ \partial_3 H(Z_{t_{j-1}},\theta_3)+2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\end{pmatrix}\Bigg],
\end{aligned}$$

$$\begin{aligned}
&\widetilde{\Psi}_3(\theta_1,\theta_2,\theta_3,\theta_1',\theta_2',\theta_3')\\
=\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\widetilde{\mathcal{D}}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\otimes\begin{pmatrix}0\\ \partial_3 H(Z_{t_{j-1}},\theta_3)+2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3)\end{pmatrix}\Bigg].
\end{aligned}$$

Let

$$M_n^{(1)} = \frac{1}{2}n^{-1/2}\sum_{j=1}^{n}\big(S^{-1}(\partial_1 S)S^{-1}\big)(Z_{t_{j-1}},\theta_1^*,\theta_3^*) \cdot\big[\widetilde{\mathcal{D}}_j(\theta_1^*,\theta_2^*,\theta_3^*)^{\otimes 2} - S(Z_{t_{j-1}},\theta_1^*,\theta_3^*)\big].$$

Let U be an open ball in $\mathbb{R}^{\mathsf{p}_2+\mathsf{p}_3}$ centered at γ^* such that $U \subset \Theta_2 \times \Theta_3$. Let $\mathcal{X}_n^{*(2,3)} = \mathcal{X}_n^{(2,3)} \cap \{\hat{\gamma}^0 \in U\}$.

Lemma 7.1. *Suppose that [A1] with $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,2,2,1,3,1)$, [A2] (i), (iii), (iv) and [A4$^\sharp$] are satisfied. Then*

$$n^{-1/2}h^{-1/2}\,\partial_2\mathbb{H}_n^{(2,3)}(\hat{\gamma}^0) = O_p(1)$$

as $n \to \infty$.

Proof. By using Lemma 5.6 and Lemma 5.5 (b) together with the convergence rate of the initial estimators, we have

$$\begin{aligned} n^{-1/2}h^{-1/2}\,\partial_2\mathbb{H}_n^{(2,3)}(\hat{\gamma}^0) &= n^{-1/2}\Psi_2(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0) \\ &= n^{-1/2}\Psi_2(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*) + O_p(1) \\ &= n^{-1/2}\widetilde{\Psi}_2(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*) + O_p(1) \end{aligned}$$

by Lemma 5.4 and Lemma 5.5 (a).

The open ball of radius r centered at θ is denoted by $U(\theta, r)$. Define the random field

$$\Phi_n^{(7.1)}(\theta) = n^{-1/2}\widetilde{\Psi}_2(\theta_1,\theta_2,\theta_3,\theta_1^*,\theta_2^*,\theta_3^*) \tag{7.1}$$

on $\theta = (\theta_1,\theta_2,\theta_3) \in U(\theta^*, r)$ for a small number r such that $U(\theta^*, r) \subset \Theta$. With the Burkholder-Davis-Gundy inequality and in particular twice differentiability of A in θ_2, we obtain

$$\sup_n \sum_{i=0,1}\sup_{\theta\in B(\theta^*,r)}\big\||\partial_\theta^i\Phi_n^{(7.1)}(\theta)|\big\|_p < \infty$$

for every $p > 1$. Therefore, Sobolev's inequality ensures

$$\sup_n \Big\|\sup_{\theta\in U(\theta^*,r)}|\Phi_n^{(7.1)}(\theta)|\Big\|_p < \infty$$

Consequently,

$$\Phi_n^{(7.1)}(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0)1_{\{(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0)\in U(\theta^*,r)\}} = O_p(1).$$

This completes the proof. □

Lemma 7.2. *Suppose that [A1] with $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,3,2)$, [A2] (i), (iii), (iv) and [A4$^\sharp$] are satisfied. Then*

$$n^{-1/2}h^{1/2}\,\partial_3\mathbb{H}_n^{(2,3)}(\hat{\gamma}^0) = O_p(1)$$

as $n \to \infty$.

Proof. The proof is similar to that of Lemma 7.1. First,

$$\begin{aligned} n^{-1/2}h^{1/2}\,\partial_3\mathbb{H}_n^{(2,3)}(\hat{\gamma}^0) &= n^{-1/2}\Psi_3(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0)\\ &= n^{-1/2}\widetilde{\Psi}_3(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*)+O_p(1). \end{aligned}$$

Then we can show the lemma in the same fashion as Lemma 7.1 with a random field. □

Let

$$\begin{aligned} B_n &= U\big(\theta_1^*, n^{-1/2}\log(nh)\big)\times U\big(\theta_2^*,(nh)^{-1/2}\log(nh)\big)\\ &\quad \times U\big(\theta_3^*, n^{-1/2}h^{1/2}\log(nh)\big), \end{aligned}$$

$$B_n' = U\big(\theta_2^*,(nh)^{-1/2}\log(nh)\big)\times U\big(\theta_3^*, n^{-1/2}h^{1/2}\log(nh)\big)$$

and

$$B_n'' = U\big(\theta_1^*, n^{-1/2}\log(nh)\big)\times U\big(\theta_3^*, n^{-1/2}h^{1/2}\log(nh)\big).$$

We will use the following random fields.

$$\begin{aligned} &\Phi_{22,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3')\\ &= -\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\begin{pmatrix} \partial_2 A(Z_{t_{j-1}},\theta_2') \\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\theta_1',\theta_2',\theta_3') \end{pmatrix}^{\otimes 2}\right], \end{aligned}$$

$$\begin{aligned} &\Phi_{22,2}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_2'',\theta_3'')\\ &= \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\ &\qquad \otimes\begin{pmatrix} \partial_2^2 A(Z_{t_{j-1}},\theta_2'') \\ 2^{-1}H_x(Z_{t_{j-1}},\theta_3'')\big[\partial_2^2 A(Z_{t_{j-1}},\theta_2'')\big] \end{pmatrix}\Bigg], \end{aligned}$$

$$\begin{aligned} &\widetilde{\Phi}_{22,2}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_2'',\theta_3'')\\ &= \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\widetilde{\mathcal{D}}_j(\theta_1',\theta_2',\theta_3')\\ &\qquad \otimes\begin{pmatrix} \partial_2^2 A(Z_{t_{j-1}},\theta_2'') \\ 2^{-1}H_x(Z_{t_{j-1}},\theta_3'')\big[\partial_2^2 A(Z_{t_{j-1}},\theta_2'')\big] \end{pmatrix}\Bigg], \end{aligned}$$

$$\begin{aligned} &\Phi_{23,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_2'',\theta_3'')\\ &= -\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\begin{pmatrix} 0 \\ 2^{-1}\partial_3 L_H(Z_{t_{j-1}},\theta_1',\theta_2',\theta_3') \end{pmatrix}\right.\\ &\qquad \left.\otimes\begin{pmatrix} \partial_2 A(Z_{t_{j-1}},\theta_2'') \\ 2^{-1}H_x(Z_{t_{j-1}},\theta_3'')\big[\partial_2 A(Z_{t_{j-1}},\theta_2'')\big] \end{pmatrix}\right], \end{aligned}$$

$$\begin{aligned}
&\Phi_{23,2}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_2'',\theta_3'')\\
&= \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\otimes\begin{pmatrix} 0 \\ 2^{-1}\partial_3 H_x(Z_{t_{j-1}},\theta_3'')\big[\partial_2 A(Z_{t_{j-1}},\theta_2'')\big]\end{pmatrix}\Bigg],
\end{aligned}$$

$$\begin{aligned}
&\Phi_{33,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3')\\
&= -\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\begin{pmatrix} 0 \\ \left\{\begin{array}{c}\partial_3 H(Z_{t_{j-1}},\theta_3') \\ +2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1',\theta_2',\theta_3')\end{array}\right\}\end{pmatrix}^{\otimes 2}\right],
\end{aligned}$$

$$\begin{aligned}
&\Phi_{33,2}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_1'',\theta_2'',\theta_3'')\\
&= \sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\Bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\
&\qquad\otimes\begin{pmatrix} 0 \\ \partial_3^2 H(Z_{t_{j-1}},\theta_3'')+2^{-1}h\partial_3^2 L_H(Z_{t_{j-1}},\theta_1'',\theta_2'',\theta_3'')\end{pmatrix}\Bigg].
\end{aligned}$$

Lemma 7.3. *Suppose that [A1] with $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,3,2,1,3,1)$, [A2] and [A4$^\sharp$] are satisfied. Then*

$$\sup_{(\theta_2,\theta_3)\in B_n'}\left|n^{-1}h^{-1}\,\partial_2^2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3)+\Gamma_{22}\right|\to^p 0$$

as $n\to\infty$.

Proof. We have

$$\begin{aligned}
n^{-1}h^{-1}\,\partial_2^2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3) &= n^{-1}\Phi_{22,1}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3)\\
&\quad +n^{-1}h^{-1/2}\Phi_{22,2}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3,\theta_2,\theta_3)
\end{aligned}\tag{7.2}$$

Apply Lemma 5.6 and Lemma 5.5 (b) to obtain

$$\begin{aligned}
&\sup_{(\theta_1,\theta_3)\in B_n''}\ \sup_{(\theta_1',\theta_2',\theta_3')\in B_n}\ \sup_{(\theta_2'',\theta_3'')\in B_n'}\Big|n^{-1}h^{-1/2}\Phi_{22,2}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_2'',\theta_3'')\\
&\qquad -n^{-1}h^{-1/2}\Phi_{22,2}(\theta_1,\theta_3,\theta_1^*,\theta_2^*,\theta_3^*,\theta_2'',\theta_3'')\Big|\\
&= o_p(1).
\end{aligned}\tag{7.3}$$

Here we used the assumption that the functions are bounded by a polynomial in z uniformly in the parameters, and the count

$$n^{-1}h^{-1/2}\times n\times h^{-1/2}\times n^{-1/2}h^{1/2}\log(nh)=\frac{\log(nh)}{\sqrt{nh}}$$

to estimate the error when replacing θ_3' by θ_3^*, as well a similar count when replacing (θ_1', θ_2') by (θ_1^*, θ_2^*).

We apply Lemmas 5.4 and 5.5 (a) to obtain

$$\begin{aligned}&\sup_{(\theta_1,\theta_3)\in B_n''}\sup_{(\theta_2'',\theta_3'')\in B_n'}\big|n^{-1}h^{-1/2}\Phi_{22,2}(\theta_1,\theta_3,\theta_1^*,\theta_2^*,\theta_3^*,\theta_2'',\theta_3'')\\ &\qquad -n^{-1}h^{-1/2}\widetilde{\Phi}_{22,2}(\theta_1,\theta_3,\theta_1^*,\theta_2^*,\theta_3^*,\theta_2'',\theta_3'')\big|\\ &= O_p\big((nh)^{-1/2}\log(nh)\big) = o_p(1).\end{aligned} \tag{7.4}$$

Since $\widetilde{\mathcal{D}}_j(\theta_1^*, \theta_2^*, \theta_3^*)$ in $\widetilde{\Phi}_{22,2}$ are martingale differences with respect to a suitable filtration, we can conclude by the random field argument with the Sobolev space of index $(1, p)$, $p > 1$, that

$$\begin{aligned}&\sup_{(\theta_1,\theta_3)\in B_n''}\sup_{(\theta_2'',\theta_3'')\in B_n'}\big|n^{-1}h^{-1/2}\widetilde{\Phi}_{22,2}(\theta_1,\theta_3,\theta_1^*,\theta_2^*,\theta_3^*,\theta_2'',\theta_3'')\big|\\ &= O_p((nh)^{-1/2}) = o_p(1)\end{aligned} \tag{7.5}$$

On the other hand,

$$\begin{aligned}&\sup_{(\theta_1,\theta_3)\in B_n''}\sup_{(\theta_1',\theta_2',\theta_3')\in B_n}\Big|n^{-1}\Phi_{22,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3')\\ &\qquad -n^{-1}\Phi_{22,1}(\theta_1^*,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)\Big| = o_p(1)\end{aligned} \tag{7.6}$$

From (7.2)-(7.6) and $[A4^\sharp]$ (i), (iii), we obtain

$$\sup_{(\theta_2,\theta_3)\in B_n'}\big|n^{-1}h^{-1}\,\partial_2^2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3)-n^{-1}\Phi_{22,1}(\theta_1^*,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)\big| = o_p(1). \tag{7.7}$$

Now the assertion of the lemma is easy to obtain if one uses $[A1]$, $[A2]$ and Lemma 5.1. □

Let

$$\begin{aligned}i(z,\theta) &= \begin{pmatrix}\partial_2 A(z,\theta_2)^\star & 2^{-1}\partial_2 L_H(z,\theta_1,\theta_2,\theta_3)^\star\\ O & \partial_3 H(z,\theta_3)^\star\end{pmatrix} S(z,\theta_1,\theta_3)^{-1}\\ &\quad\times\begin{pmatrix}\partial_2 A(z,\theta_2) & O\\ 2^{-1}\partial_2 L_H(z,\theta_1,\theta_2,\theta_3) & \partial_3 H(z,\theta_3)\end{pmatrix}.\end{aligned} \tag{7.8}$$

Then simple calculus with (3.3) and

$$\partial_2 L_H(z,\theta_1,\theta_2,\theta_3) = H_x(z,\theta_3)\big[\partial_2 A(z,\theta_2)\big]$$

yield

$$i(z,\theta) = \mathrm{diag}\big[i(z,\theta)_{22}, i(z,\theta)_{33}\big], \tag{7.9}$$

where

$$i(z,\theta)_{22} = \partial_2 A(z,\theta_2)^\star C(z,\theta_1)^{-1}\partial_2 A(z,\theta_2)$$

and

$$i(z,\theta)_{33} = 12\partial_3 H(z,\theta_3)^\star V(z,\theta_1,\theta_3)^{-1}\partial_3 H(z,\theta_3).$$

Lemma 7.4. *Suppose that [A1] with $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,3,1)$, [A2] and [A4$^\sharp$] are satisfied. Then*

$$\sup_{(\theta_2,\theta_3)\in B'_n} \big|n^{-1}\,\partial_3\partial_2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3)\big| \to^p 0$$

as $n\to\infty$.

Proof. From (7.8), we see

$$S(z,\theta_1,\theta_3)^{-1}\left[\begin{pmatrix} 0 \\ \partial_3 H(z,\theta_3)\end{pmatrix}, \begin{pmatrix} \partial_2 A(z,\theta_2) \\ 2^{-1}\partial_2 L_H(z,\theta_1,\theta_2,\theta_3)\end{pmatrix}\right] = 0.$$

Then,

$$\begin{aligned} n^{-1}\,\partial_3\partial_2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3) &= n^{-1}h\Phi_{23,1}(\hat\theta_1^0,\theta_3,\hat\theta_1^0,\theta_2,\theta_3,\theta_2,\theta_3)+\mathfrak{e}_n(\theta_2,\theta_3) \\ &\quad +n^{-1}h^{1/2}\Phi_{23,2}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3,\theta_2,\theta_3)\end{aligned}$$

where

$$\sup_{(\theta_2,\theta_3)\in B'_n}\big|\mathfrak{e}_n(\theta_2,\theta_3)\big| = O_p(n^{-1/2}h^{1/2})$$

as $n\to\infty$. Indeed, first replace $\hat S(Z_{t_{j-1}},\hat\theta_3^0)^{-1}$ by $\hat S(Z_{t_{j-1}},\theta_3)^{-1}$, and next use the above equality to remove the term $\partial_3 H(Z_{t_{j-1}},\theta_3)$. Now it is not difficult to show the desired result. □

Lemma 7.5. *Suppose that [A1] with $(i_A, j_A, i_B, j_B, i_H, j_H) = (1,1,2,1,3,2)$, [A2] and [A4$^\sharp$] are satisfied. Then*

$$\sup_{(\theta_2,\theta_3)\in B'_n} \big|n^{-1}h\,\partial_3^2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3)+\Gamma_{33}\big| \to^p 0$$

as $n\to\infty$.

Proof. By definition,

$$\begin{aligned} n^{-1}h\,\partial_3^2\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3) &= n^{-1}\Phi_{33,1}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3) \\ &\quad +n^{-1}h^{1/2}\Phi_{33,2}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3,\hat\theta_1^0,\theta_2,\theta_3).\end{aligned}$$

$\Phi_{33,1}$ involves the first derivative ∂_3, and $\Phi_{33,2}$ does the second derivative ∂_3^2. First applying Lemma 5.6 and Lemma 5.5 (b), and next Lemma 5.3 (b), we have

$$\begin{aligned}
&\sup_{(\theta_2,\theta_3)\in B'_n}\left|n^{-1}h^{1/2}\Phi_{33,2}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3,\hat\theta_1^0,\theta_2,\theta_3)\right|\\
\leq\;& \sup_{(\theta_2,\theta_3)\in B'_n}\left|n^{-1}h^{1/2}\Phi_{33,2}(\hat\theta_1^0,\hat\theta_3^0,\theta_1^*,\theta_2^*,\theta_3^*,\hat\theta_1^0,\theta_2,\theta_3)\right|\\
&+O_p(n^{-1/2}h^{1/2}\log(nh))\\
=\;& O_p(h^{1/2}).
\end{aligned}$$

Moreover, it is easy to show

$$\sup_{(\theta_2,\theta_3)\in B'_n}\left|n^{-1}\Phi_{33,1}(\hat\theta_1^0,\hat\theta_3^0,\hat\theta_1^0,\theta_2,\theta_3)+\Gamma_{33}\right|\to^p 0$$

from [$A1$], [$A2$] and [$A4^\sharp$] with the aid of Lemma 5.1. □

Let

$$a_n=\begin{pmatrix} n^{-1/2}h^{-1/2} & 0\\ 0 & n^{-1/2}h^{1/2}\end{pmatrix}.$$

Lemma 7.6. *Suppose that [$A1$] with $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,3,2,1,3,2)$, [$A2$] and [$A4^\sharp$] are satisfied. Then*

$$\sup_{(\theta_2,\theta_3)\in B'_n}\left|a_n\partial^2_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\theta_2,\theta_3)a_n+\Gamma^{(2,3)}(\theta^*)\right|\;\to^p\;0 \tag{7.10}$$

where

$$\Gamma^{(2,3)}(\theta^*)=\begin{pmatrix}\Gamma_{22} & O\\ O & \Gamma_{33}\end{pmatrix}.$$

Proof. The convergence (7.10) follows from Lemmas 7.3, 7.4 and 7.5. □

Lemma 7.7. *Suppose that [$A1$] with $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,3,2,1,3,2)$, [$A2$] and [$A4^\sharp$] are satisfied. If $\Gamma^{(2,3)}(\theta^*)$ is invertible, then $P[\mathcal{X}_n^{*(2,3)}]\to 1$ as $n\to\infty$.*

Proof. By Lemmas 7.1 and 7.2,

$$a_n\partial_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\hat\gamma^0)=O_p(1)$$

and by Lemma 7.6,

$$\left(a_n\partial^2_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\hat\gamma^0)a_n\right)^{-1}=O_p(1).$$

Therefore,

$$\left(\partial^2_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\hat\gamma^0)\right)^{-1}\partial_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\hat\gamma^0)=O_p((nh)^{-1/2})$$

as $n\to\infty$. This means $P[\mathcal{X}_n^{*(2,3)}]\to 1$. □

Let

$$\begin{aligned} M_n^{(2)} &= n^{-1/2}\sum_{j=1}^{n} S(Z_{t_{j-1}},\theta_1^*,\theta_3^*)^{-1}\Bigg[\begin{pmatrix} h^{-1/2}B(Z_{t_{j-1}},\theta_1^*)\Delta_j w \\ h^{-3/2}\kappa(Z_{t_{j-1}},\theta_1^*,\theta_3^*)\zeta_j \end{pmatrix} \\ &\qquad\qquad \otimes \begin{pmatrix} \partial_2 A(Z_{t_{j-1}},\theta_2^*) \\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\theta_1^*,\theta_2^*,\theta_3^*) \end{pmatrix}\Bigg] \\ &= n^{-1/2}\sum_{j=1}^{n} C(Z_{t_{j-1}},\theta_1^*)^{-1}\big[h^{-1/2}B(Z_{t_{j-1}},\theta_1^*)\Delta_j w,\ \partial_2 A(Z_{t_{j-1}},\theta_2^*)\big]. \end{aligned} \tag{7.11}$$

Lemma 7.8. *Suppose that [A1] with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,3,1)$, *[A2] and [A4$^\sharp$] are satisfied. Then*

$$n^{-1/2}h^{-1/2}\,\partial_2\mathbb{H}_n^{(2,3)}(\theta_2^*,\theta_3^*) - M_n^{(2)} \to^p 0$$

as $n\to\infty$.

Proof. By using Lemma 5.5 (b) together with the convergence rate of the estimators $\hat\theta_1^0$ and $\hat\theta_3^0$, and next by Lemma 5.5 (a) and Lemma 5.4, we have

$$\begin{aligned} & n^{-1/2}h^{-1/2}\,\partial_2\mathbb{H}_n^{(2,3)}(\theta_2^*,\theta_3^*) \\ =\ & n^{-1/2}\sum_{j=1}^{n}\hat S(Z_{t_{j-1}},\hat\theta_3^0)^{-1}\Bigg[\mathcal{D}_j(\hat\theta_1^0,\theta_2^*,\theta_3^*) \\ &\qquad\qquad \otimes\begin{pmatrix} \partial_2 A(Z_{t_{j-1}},\theta_2^*) \\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\hat\theta_1^0,\theta_2^*,\theta_3^*) \end{pmatrix}\Bigg] \\ =\ & n^{-1/2}\sum_{j=1}^{n}\hat S(Z_{t_{j-1}},\hat\theta_3^0)^{-1}\Bigg[\mathcal{D}_j(\theta_1^*,\theta_2^*,\theta_3^*) \\ &\qquad\qquad \otimes\begin{pmatrix} \partial_2 A(Z_{t_{j-1}},\theta_2^*) \\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\hat\theta_1^0,\theta_2^*,\theta_3^*) \end{pmatrix}\Bigg] \\ & +O_p(h^{1/2}) \\ =\ & n^{-1/2}\sum_{j=1}^{n}\hat S(Z_{t_{j-1}},\theta_3^*)^{-1}\Bigg[\widetilde{\mathcal{D}}_j(\theta_1^*,\theta_2^*,\theta_3^*) \\ &\qquad\qquad \otimes\begin{pmatrix} \partial_2 A(Z_{t_{j-1}},\theta_2^*) \\ 2^{-1}\partial_2 L_H(Z_{t_{j-1}},\hat\theta_1^0,\theta_2^*,\theta_3^*) \end{pmatrix}\Bigg] \\ & +O_p(\sqrt{n}h)+O_p(h^{1/2}). \end{aligned} \tag{7.12}$$

Here we used the derivative $\partial_1 H$.

We consider the random field

$$\Phi_n^{(7.13)}(u_1) \quad = \quad n^{-1/2}\widetilde{\Psi}_2(\theta_1(u_1),\theta_2^*,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*) \tag{7.13}$$

on $\{u_1 \in \mathbb{R}^{\mathsf{p}_1};\ |u_1|<1\}$, where $\theta_1(u_1)=\theta_1^*+n^{-1/2}(\log n)u_1$. Then L^p-estimate of

$$\partial_1^i\{\Phi_n^{(7.13)}(u_1)-\Phi_n^{(7.13)}(0)\} \quad (i=0,1)$$

yields

$$\sup_{u_1\in U(0,1)}\left|\Phi_n^{(7.13)}(u_1)-\Phi_n^{(7.13)}(0)\right| \to^p 0,$$

in particular,

$$\Phi_n^{(7.13)}(u_1^\dagger)-\Phi_n^{(7.13)}(0)\to^p 0$$

where $u_1^\dagger = n^{1/2}(\log n)^{-1}(\hat{\theta}_1-\theta_1^*)$. Obviously, $M_n^{(2)}-\Phi_n^{(7.13)}(0)\to^p 0$. Since the first term on the right-hand side of (7.12) is nothing but $\Phi_n^{(7.13)}(u_1^\dagger)$ on an event the probability of which goes to 1, we have already obtained the result. □

Lemma 7.9. *Suppose that* [A1] *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,3,1)$, [A2] *and* [A4] *are satisfied. Then*

$$n^{-1/2}h^{1/2}\,\partial_3\mathbb{H}_n^{(2,3)}(\theta_2^*,\theta_3^*)-M_n^{(3)}\to^p 0$$

as $n\to\infty$.

Proof. We have

$$\begin{aligned} & n^{-1/2}h^{1/2}\,\partial_3\mathbb{H}_n^{(2,3)}(\theta_2^*,\theta_3^*) \\ = \quad & n^{-1/2}\sum_{j=1}^n S(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_3^0)^{-1}\Bigg[\mathcal{D}_j(\hat{\theta}_1^0,\theta_2^*,\theta_3^*) \\ & \qquad \otimes\begin{pmatrix} 0 \\ \partial_3 H(Z_{t_{j-1}},\theta_3^*)+2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_2^*,\theta_3^*)\end{pmatrix}\Bigg]. \end{aligned}$$

Then this lemma can be proved in the same way as Lemma 7.8. □

Let

$$M_n^{(2,3)}=\begin{pmatrix} M_n^{(2)} \\ M_n^{(3)}\end{pmatrix}.$$

Combining Lemmas 7.8 and 7.9, with the identity in the proof of Lemma 7.4, we obtain the following lemma.

Lemma 7.10. *Suppose that* [A1] *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,3,1)$, [A2] *and* [A4$^\sharp$] *are satisfied. Then*

$$a_n\partial_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\theta_2^*,\theta_3^*)-M_n^{(2,3)}\to^p 0$$

and $M_n^{(2,3)}\to^d N(0,\Gamma^{(2,3)}(\theta^*))$ *as* $n\to\infty$. *In particular,*

$$a_n\partial_{(\theta_2,\theta_3)}\mathbb{H}_n^{(2,3)}(\theta_2^*,\theta_3^*)\to^d N(0,\Gamma^{(2,3)}(\theta^*))$$

as $n\to\infty$.

Theorem 7.11. *Suppose that* [A1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1, 3, 2, 1, 3, 2)$, [A2] *and* [A4$^\sharp$] *are satisfied. If* $\Gamma^{(2,3)}(\theta^*)$ *is invertible, then*

$$a_n^{-1}(\hat{\gamma} - \gamma^*) - (\Gamma^{(2,3)}(\theta^*))^{-1} M_n^{(2,3)} \to^p 0 \tag{7.14}$$

as $n \to \infty$*. In particular,*

$$a_n^{-1}(\hat{\gamma} - \gamma^*) \to^d N(0, (\Gamma^{(2,3)}(\theta^*))^{-1}) \tag{7.15}$$

as $n \to \infty$*.*

Proof. Let

$$\begin{aligned} \mathcal{X}_n^{**(2,3)} &= \mathcal{X}_n^{*(2,3)} \cap \big\{(\hat{\theta}_1^0, \hat{\gamma}^0) \in B_n\big\} \\ &\quad \cap \Big\{ \sup_{\gamma \in B_n'} \big| a_n \partial^2_{(\theta_2,\theta_3)} \mathbb{H}_n^{(2,3)}(\gamma) a_n + \Gamma^{(2,3)}(\theta^*) \big| < c \Big\}. \end{aligned}$$

Here c is a postive constant and we will make it sufficiently small. Then

$$P[\mathcal{X}_n^{**(2,3)}] \to 1$$

thanks to Lemmas 7.7 and 7.6. On the event $\mathcal{X}_n^{**(2,3)}$, we apply Taylor's formula to obtain

$$\begin{aligned} & a_n^{-1}(\hat{\gamma} - \gamma^*) \\ = \ & \big[a_n \partial^2_{(\theta_2,\theta_3)} \mathbb{H}_n^{(2,3)}(\hat{\gamma}^0) a_n\big]^{-1} \Big\{ - a_n \partial_{(\theta_2,\theta_3)} \mathbb{H}_n^{(2,3)}(\gamma^*) \\ & + a_n \int_0^1 \big[\partial^2_{(\theta_2,\theta_3)} \mathbb{H}_n^{(2,3)}(\hat{\gamma}^0) - \partial^2_{(\theta_2,\theta_3)} \mathbb{H}_n^{(2,3)}(\hat{\gamma}(u))\big] du a_n \, a_n^{-1}\big(\hat{\gamma}^0 - \gamma^*\big) \Big\} \end{aligned}$$

where $\hat{\gamma}(u) = \gamma^* + u(\hat{\gamma}^0 - \gamma^*)$. Then Lemmas 7.6 and 7.10 give (7.14). Then the martingale central limit theorem gives (7.15). □

The following notation for random fields will be used.

$$\begin{aligned} & \Psi_{1,1}(\theta_1, \theta_2, \theta_3, \theta_1', \theta_2', \theta_3') \\ = \ & \sum_{j=1}^n S(Z_{t_{j-1}}, \theta_1, \theta_3)^{-1} \left[\mathcal{D}_j(\theta_1', \theta_2', \theta_3'), \begin{pmatrix} 0 \\ 2^{-1} \partial_1 L_H(Z_{t_{j-1}}, \theta_1, \theta_2, \theta_3) \end{pmatrix} \right] \\ = \ & \sum_{j=1}^n S(Z_{t_{j-1}}, \theta_1, \theta_3)^{-1} \Bigg[\mathcal{D}_j(\theta_1', \theta_2', \theta_3') \\ & \qquad \otimes \begin{pmatrix} 0 \\ 4^{-1} H_{xx}(z, \theta_3)[\partial_1 C(Z_{t_{j-1}}, \theta_1)] \end{pmatrix} \Bigg], \end{aligned}$$

$$\begin{aligned} & \Psi_{1,2}(\theta_1, \theta_3, \theta_1', \theta_2', \theta_3') \\ = \ & \frac{1}{2} \sum_{j=1}^n \big(S^{-1}(\partial_1 S)) S^{-1}\big)(Z_{t_{j-1}}, \theta_1, \theta_3) \big[\mathcal{D}_j(\theta_1', \theta_2', \theta_3')^{\otimes 2} - S(Z_{t_{j-1}}, \theta_1, \theta_3)\big], \end{aligned}$$

$$\begin{aligned}&\Psi_{11,1}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3')\\ =\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\begin{pmatrix}0\\ 2^{-1}\partial_1 L_H(Z_{t_{j-1}},\theta_1',\theta_2',\theta_3')\end{pmatrix}^{\otimes 2}\right],\end{aligned}$$

$$\begin{aligned}&\Psi_{11,2}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_1'',\theta_2'',\theta_3'')\\ =\ &\sum_{j=1}^n S(Z_{t_{j-1}},\theta_1,\theta_3)^{-1}\left[\mathcal{D}_j(\theta_1',\theta_2',\theta_3'),\ \begin{pmatrix}0\\ 2^{-1}\partial_1^2 L_H(Z_{t_{j-1}},\theta_1'',\theta_2'',\theta_3'')\end{pmatrix}\right],\end{aligned}$$

$$\begin{aligned}&\Psi_{11,3}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3')\\ =\ &\sum_{j=1}^n \partial_1\big\{S^{-1}(\partial_1 S)S^{-1}(Z_{t_{j-1}},\theta_1,\theta_3)\big\}\big[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')^{\otimes 2}-S(Z_{t_{j-1}},\theta_1',\theta_3')\big],\end{aligned}$$

$$\Psi_{11,4}(\theta_1,\theta_3)=\sum_{j=1}^n (S^{-1}(\partial_1 S)S^{-1})(Z_{t_{j-1}},\theta_1,\theta_3)\big[\partial_1 S(Z_{t_{j-1}},\theta_1,\theta_3)\big],$$

$$\begin{aligned}&\Psi_{11,5}(\theta_1,\theta_3,\theta_1',\theta_2',\theta_3',\theta_1'',\theta_2'',\theta_3'')\\ =\ &2\sum_{j=1}^n (S^{-1}(\partial_1 S)S^{-1})(Z_{t_{j-1}},\theta_1,\theta_3)\bigg[\mathcal{D}_j(\theta_1',\theta_2',\theta_3')\\ &\qquad\otimes\begin{pmatrix}0\\ 2^{-1}\partial_1 L_H(Z_{t_{j-1}},\theta_1'',\theta_2'',\theta_3'')\end{pmatrix}\bigg].\end{aligned}$$

We sometimes keep parameters in notation even when some of them do not appear in a specific expression of the formula, if such an expression is not necessary for later use; e.g.. $\Psi_{1,1}$ does not depend on θ_2 in fact.

Lemma 7.12. *Suppose that* [A1] *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,3,3,1)$, [A2] *and* [A4$^\sharp$] *are satisfied. Then, for any sequence of positive numbers* r_n *tending to* 0,

$$\sup_{\theta_1\in U(\theta_1^*,r_n)}\left|n^{-1}\,\partial_1^2\mathbb{H}_n^{(1)}(\theta_1)+\Gamma_{11}\right|\to^p 0 \tag{7.16}$$

as $n\to\infty$.

Proof. By definition,

$$\begin{aligned}n^{-1}\,\partial_1^2\mathbb{H}_n^{(1)}(\theta_1)\ =\ &-n^{-1}h\Psi_{11,1}(\theta_1,\hat\theta_3^0,\theta_1,\hat\theta_2^0,\hat\theta_3^0)\\ &+n^{-1}h^{1/2}\Psi_{11,2}(\theta_1,\hat\theta_3^0,\theta_1,\hat\theta_2^0,\hat\theta_3^0,\theta_1,\hat\theta_2^0,\hat\theta_3^0)\\ &+\frac{1}{2}n^{-1}\Psi_{11,3}(\theta_1,\hat\theta_3^0,\theta_1,\hat\theta_2^0,\hat\theta_3^0)\\ &-\frac{1}{2}n^{-1}\Psi_{11,4}(\theta_1,\hat\theta_3^0)\quad\text{(this term will remain)}\\ &-n^{-1}h^{1/2}\Psi_{11,5}(\theta_1,\hat\theta_3^0,\theta_1,\hat\theta_2^0,\hat\theta_3^0,\theta_1,\hat\theta_2^0,\hat\theta_3^0)\end{aligned}$$

We will use Condition $[A4^\sharp]$ for $\hat{\theta}_2^0$ and $\hat{\theta}_3^0$, and the estimate $|\theta_1-\theta_1^*|<r_n$ for $\theta_1\in U(\theta_1^*,r_n)$. Then

$$\begin{aligned}
&\sup_{\theta_1\in U(\theta_1^*,r_n)}\left|n^{-1}\,\partial_1^2\mathbb{H}_n^{(1)}(\theta_1)+\Gamma_{11}\right|\\
\le\;& O_p(h)\\
&+n^{-1}h^{1/2}\sup_{\theta_1\in U(\theta_1^*,r_n)}\left|\Psi_{11,2}(\theta_1,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*,\theta_1,\hat{\theta}_2^0,\hat{\theta}_3^0)\right|\\
&\qquad\qquad+h^{1/2}O_p(n^{-1/2}+h^{1/2})\quad(\text{Lemmas } 5.6 \text{ and } 5.5(b))\\
&+n^{-1}\sup_{\theta_1\in U(\theta_1^*,r_n)}\left|\Psi_{11,3}(\theta_1,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*)\right|+O_p(h^{1/2}+n^{-1/2})+O_p(r_n)\\
&\quad(\text{Lemmas } 5.6 \text{ and } 5.5(b))\\
&+\left|-\frac{1}{2}n^{-1}\Psi_{11,4}(\theta_1^*,\theta_3^*)+\Gamma_{11}\right|+O_p(r_n)+O_p(n^{-1/2}h^{1/2})\\
&+n^{-1}h^{1/2}\sup_{\theta_1\in U(\theta_1^*,r_n)}\left|\Psi_{11,5}(\theta_1,\hat{\theta}_3^0,\theta_1,\theta_2^*,\theta_3^*,\theta_1,\hat{\theta}_2^0,\hat{\theta}_3^0)\right|\\
&\qquad\qquad+h^{1/2}O_p(h^{1/2}+n^{-1/2})\quad(\text{Lemmas } 5.6 \text{ and } 5.5(b))\\
=\;& O_p(h)\\
&+O_p(h^{1/2})\quad(\text{Lemma } 5.3(b))\\
&+O_p(h^{1/2})+O_p(n^{-1/2})+O_p(r_n)\\
&\quad(\text{random field argument with orthogonality})\\
&+o_p(1)\quad(\text{Lemma } 5.1(a))\\
&+O_p(h^{1/2})\quad(\text{Lemma } 5.5(b))\\
=\;& o_p(1)
\end{aligned}$$

We remark that the used lemmas and appearing functions here require the regularity indices $(i_A,j_A,i_B,j_B,i_H,j_H)$ for $[A1]$ as follows: $(1,0,1,0,3,0)$ for Lemma 5.3(b); $(1,1,2,1,2,0)$ for Lemma 5.5(b); $(0,0,0,0,2,1)$ for Lemma 5.6; $j_B=3$, $j_H=1$ for random field argument for $\Psi_{11,3}$. □

Lemma 7.13. *Suppose that* $[A1]$ *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,2,1)$, $[A2]$ *and* $[A4^\sharp]$ *are satisfied. Then*

$$n^{-1/2}\partial_1\mathbb{H}_n^{(1)}(\hat{\theta}_1^0)=O_p(1)\tag{7.17}$$

as $n\to\infty$.

Proof. We have the expression

$$\begin{aligned}
n^{-1/2}\partial_1\mathbb{H}_n^{(1)}(\hat{\theta}_1^0)\;=\;& n^{-1/2}h^{1/2}\Psi_{1,1}(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0)\\
&+n^{-1/2}\Psi_{1,2}(\hat{\theta}_1^0,\hat{\theta}_3^0,\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0).
\end{aligned}$$

We use $[A4^\sharp]$ together with Lemmas 5.6 and 5.5 (b) to show

$$n^{-1/2}h^{1/2}\Psi_{1,1}(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0)$$

$$
\begin{aligned}
&= n^{-1/2}h^{1/2}\Psi_{1,1}(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*)+o_p(1)\\
&= o_p(1) \;=\; O_p(1)
\end{aligned}
$$

and

$$
\begin{aligned}
n^{-1/2}\Psi_{1,2}(\hat{\theta}_1^0,\hat{\theta}_3^0,\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_3^0) &= n^{-1/2}\Psi_{1,2}(\hat{\theta}_1^0,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*)+O_p(1)\\
&= O_p(1)
\end{aligned}
$$

as $n\to\infty$. Here random field argument was used. □

Lemma 7.14. *Suppose that* $[A1]$ *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,3,1)$, $[A2]$ *and* $[A4^\sharp]$ *are satisfied. Then*

$$
n^{-1/2}\partial_1\mathbb{H}_n^{(1)}(\theta_1^*)-M_n^{(1)}\to^p 0 \tag{7.18}
$$

as $n\to\infty$. *In particular,*

$$
n^{-1/2}\partial_1\mathbb{H}_n^{(1)}(\theta_1^*)\to^d N\big(0,\Gamma_{11}\big) \tag{7.19}
$$

as $n\to\infty$.

Proof. We have

$$
\begin{aligned}
&\mathsf{E}_j(\theta_2,\theta_3)\\
:=\;& \mathcal{D}_j(\theta_1^*,\theta_2,\theta_3)-\mathcal{D}_j(\theta_1^*,\theta_2^*,\theta_3^*)\\
=\;& \left(\begin{array}{c} h^{1/2}\big(A(Z_{t_{j-1}},\theta_2^*)-A(Z_{t_{j-1}},\theta_2)\big)\\ \left\{\begin{array}{c} h^{-1/2}\big(H(Z_{t_{j-1}},\theta_3^*)-H(Z_{t_{j-1}},\theta_3)\big)\\ +2^{-1}h^{1/2}\big(L_H(Z_{t_{j-1}},\theta_1^*,\theta_2^*,\theta_3^*)-L_H(Z_{t_{j-1}},\theta_1^*,\theta_2,\theta_3)\big)\end{array}\right\}\end{array}\right).
\end{aligned}
$$

Define the random field $\Xi_n(u_2,u_3)$ on $(u_2,u_3)\in U(0,1)^2$ by

$$
\begin{aligned}
&\Xi_n(u_2,u_3)\\
=\;& n^{-1/2}\sum_{j=1}^n\big(S^{-1}(\partial_1S)S^{-1}\big)(Z_{t_{j-1}},\theta_1^*,\theta_3^*+r_n^{(3)}u_3)\\
&\qquad\cdot\Big[\mathcal{D}_j(\theta_1^*,\theta_2^*,\theta_3^*)\otimes\mathsf{E}_j(\theta_2^*+r_n^{(2)}u_2,\theta_3^*+r_n^{(3)}u_3)\Big]
\end{aligned}
$$

where $r_n^{(2)}=(nh)^{-1/2}\log(nh)$ and $r_n^{(3)}=n^{-1/2}h^{1/2}\log(nh)$. Then the Burkholder-Davis-Gundy inequality gives

$$
\lim_{n\to\infty}\sup_{(u_2,u_3)\in U(0,1)^2}\sum_{i=0,1}\big\|\partial^i_{(u_2,u_3)}\Xi_n(u_2,u_3)\big\|_p=0,
$$

which implies

$$
\sup_{(u_2,u_3)\in U(0,1)^2}\big|\Xi_n(u_2,u_3)\big|\to^p 0,
$$

and hence under $[A4^\sharp]$,

$$n^{-1/2}\sum_{j=1}^{n}\big(S^{-1}(\partial_1 S)S^{-1}\big)(Z_{t_{j-1}},\theta_1^*,\hat{\theta}_3^0)\Big[\mathcal{D}_j(\theta_1^*,\theta_2^*,\theta_3^*)\otimes \mathsf{E}_j(\hat{\theta}_2^0,\hat{\theta}_3^0)\Big] \quad \to^p \quad 0 \tag{7.20}$$

as $n\to\infty$. It is easier to see

$$n^{-1/2}\sum_{j=1}^{n}\big(S^{-1}(\partial_1 S)S^{-1}\big)(Z_{t_{j-1}},\theta_1^*,\hat{\theta}_3^0)\big[\mathsf{E}_j(\hat{\theta}_2^0,\hat{\theta}_3^0)^{\otimes 2}\big] \to^p 0 \tag{7.21}$$

as $n\to\infty$. From (7.20) and (7.21),

$$\begin{aligned} & n^{-1/2}\Psi_{1,2}(\theta_1^*,\hat{\theta}_3^0,\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0) \\ = \; & n^{-1/2}\Psi_{1,2}(\theta_1^*,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*)+o_p(1) \\ = \; & n^{-1/2}\Psi_{1,2}(\theta_1^*,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)+o_p(1) \end{aligned} \tag{7.22}$$

as $n\to\infty$, where the last equality is by $[A4^\sharp]$.

On the other hand, by $[A4^\sharp]$ and Lemmas 5.6 and 5.5 (b), we obtain

$$\begin{aligned} & n^{-1/2}h^{1/2}\Psi_{1,1}(\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0) \\ = \; & n^{-1/2}h^{1/2}\Psi_{1,1}(\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\theta_2^*,\theta_3^*)+o_p(1). \end{aligned} \tag{7.23}$$

By random field argument applied to the first term on the right-hand side of (7.23),

$$n^{-1/2}h^{1/2}\Psi_{1,1}(\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0)=o_p(1). \tag{7.24}$$

Consequently, from (7.22) and (7.24), we obtain the convergence (7.18) since

$$\begin{aligned} n^{-1/2}\partial_1\mathbb{H}_n^{(1)}(\theta_1^*) &= n^{-1/2}h^{1/2}\Psi_{1,1}(\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0,\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0) \\ &\quad +n^{-1/2}\Psi_{1,2}(\theta_1^*,\hat{\theta}_3^0,\theta_1^*,\hat{\theta}_2^0,\hat{\theta}_3^0) \\ &= n^{-1/2}\Psi_{1,2}(\theta_1^*,\theta_3^*,\theta_1^*,\theta_2^*,\theta_3^*)+o_p(1) \\ &= M_n^{(1)}+o_p(1) \end{aligned}$$

by using Lemmas 5.4 and 5.5 (a). Convergence (7.19) follows from this fact and Lemma 5.1 with $[A2]$, □

Proof of Theorem 4.1. Let

$$\mathcal{X}_n^{***}=\mathcal{X}_n^{(1)}\cap\mathcal{X}_n^{**(2,3)}\cap\left\{\sup_{\theta_1\in B_n'''}\left|n^{-1}\partial_1^2\mathbb{H}_n^{(1)}(\theta_1)+\Gamma_{11}\right|<c_1\right\}$$

where $B_n'''=U(\theta_1^*,n^{-1/2}\log n)$, and c_1 is a sufficiently small number such that $|\mathsf{A}+\Gamma_{11}|<c_1$ implies $\det\mathsf{A}\neq 0$ for any $\mathsf{p}_1\times\mathsf{p}_1$ matrix A. We obtain $P[\mathcal{X}_n^{***}]\to 1$ from Lemmas 7.13 and 7.12 as well as the proof of Lemma 7.11.

For large n, on the event $\mathcal{X}_n^{***}$, we apply Taylor's formula to obtain

$$\begin{aligned} & n^{1/2}(\hat{\theta}_1 - \theta_1^*) \\ = \;& \big[n^{-1}\partial_{\theta_1}^2 \mathbb{H}_n^{(1)}(\hat{\theta}_1^0)\big]^{-1}\bigg\{ - n^{-1/2}\partial_{\theta_1}\mathbb{H}_n^{(1)}(\theta_1^*) \\ & + n^{-1}\int_0^1 \big[\partial_1^2\mathbb{H}_n^{(1)}(\hat{\theta}_1^0) - \partial_1^2\mathbb{H}_n^{(1)}(\hat{\theta}_1(u))\big] du\, n^{1/2}\big(\hat{\theta}_1^0 - \theta_1^*\big)\bigg\} \end{aligned}$$

where $\hat{\theta}_1(u) = \theta_1^* + u(\hat{\theta}_1^0 - \theta^*)$. Then we obtain

$$n^{1/2}\big(\hat{\theta}_1 - \theta_1^*\big) - \Gamma_{11}^{-1} M_n^{(1)} \to^p 0 \tag{7.25}$$

as $n \to \infty$ from Lemmas 7.12 and 7.14. Therefore the convergence of $b_n^{-1}(\hat{\theta} - \theta^*)$ follows from the martingale central limit theorem and the relations (7.14) and (7.25). □

8. Discussion on the estimation of θ_3 when only information of $\Delta_j Y$ is available

In Sections 3 and 4, the estimators for θ_3 used the information of $\Delta_j X$ as well as $\Delta_j Y$, given covariates $(Z_{t_j})_{j=0,\dots,n}$. Then a natural question is what occurs when only the information of $\Delta_j Y$, i.e., the martingale part of $\Delta_j Y$, is available?

It is possible to construct a QMLE $\hat{\vartheta}_3$ for θ_3 based on the quasi-log likelihood function

$$\begin{aligned} & \mathcal{H}_n^{(3)}(\theta_3) \\ = \;& -\frac{1}{2}\sum_{j=1}^n \bigg\{ 3V(Z_{t_{j-1}}, \hat{\theta}_1^0, \theta_3)^{-1}\big[\big\{h^{-3/2}\big(\Delta_j Y - hG_n(Z_{t_{j-1}}, \hat{\theta}_1^0, \hat{\theta}_2^0, \theta_3)\big)\big\}^{\otimes 2}\big] \\ & + \log\det\big(3^{-1}V(Z_{t_{j-1}}, \hat{\theta}_1^0, \theta_3)\big)\bigg\} \end{aligned}$$

with some initial estimators $\hat{\theta}_i^0$ for θ_i, $i = 1, 2$.

Let

$$\mathcal{Y}_n^{(3)}(\theta_3) = n^{-1}h\big\{\mathcal{H}_n^{(3)}(\theta_3) - \mathcal{H}_n^{(3)}(\theta_3^*)\big\}$$

and let

$$\mathcal{Y}^{(3)}(\theta_3) = -\frac{3}{2}\int V(z, \theta_1^*, \theta_3)^{-1}\big[\big(H(z, \theta_3) - H(z, \theta_3)\big)^{\otimes 2}\big]\nu(dz).$$

Consistency of $\hat{\vartheta}_3$ is obtained if the initial estimators are consistent.

Proposition 8.1. *Suppose that* [A1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1, 1, 2, 1, 3, 1)$ *and* [A2] *(i), (ii), (iv) are satisfied. Then*

$$\sup_{\theta_3 \in \overline{\Theta}_3} \big|\mathcal{Y}_n^{(3)}(\theta_3) - \mathcal{Y}^{(3)}(\theta_3)\big| \to^p 0 \tag{8.1}$$

as $n \to \infty$, *if* $\hat{\theta}_1^0 \to^p \theta_1^*$ *and* $\hat{\theta}_2^0 \to^p \theta_2^*$. *Moreover,* $\hat{\vartheta}_3 \to^p \theta_3^*$ *if* [A3] *(iii) is additionally satisfied.*

Proof. We have a decomposition of $\mathcal{Y}_n^{(3)}(\theta_3)$:

$$\begin{aligned}\mathcal{Y}_n^{(3)}(\theta_3) &= -\frac{1}{2n}\sum_{j=1}^{n} 3V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)^{-1} \\ &\qquad \cdot\Big[\big\{G_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)-G_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3^*)\big\}^{\otimes 2}\Big] \\ &\quad +R_n^{(8.3)}(\theta_3) \end{aligned} \tag{8.2}$$

where

$$\begin{aligned} R_n^{(8.3)}(\theta_3) &= \frac{h}{n}\sum_{j=1}^{n} 3V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)^{-1}\Big[h^{-3/2}\big(\Delta_j Y - hG_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3^*)\big) \\ &\qquad \otimes h^{-1/2}\big(G_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)-G_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3^*)\big)\Big] \\ &\quad -\frac{h}{2n}\sum_{j=1}^{n}\Big\{\big(3V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)^{-1}-3V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3^*)^{-1}\big) \\ &\qquad \cdot\Big[\big\{h^{-3/2}\big(\Delta_j Y - hG_n(Z_{t_{j-1}},,\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3^*)\big)\big\}^{\otimes 2}\Big] \\ &\qquad +\log\frac{\det V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)}{\det V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3^*)}\Big\} \end{aligned} \tag{8.3}$$

Applying Lemma 5.5 (a), we see the first sum on the right-hand side of (8.3) is $O_p(h^{1/2})$ and the second sum is $O_p(h)$ as $n\to\infty$. These estimates are uniform in $\theta_3\in\overline{\Theta}_3$. In particular, $\sup_{\theta_3\in\overline{\Theta}_3}\big|R_n^{(8.3)}(\theta_3)\big| = o_p(1)$. For the first sum on the right-hand side of (8.2), one can remove the terms involving $L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)$ uniformly in θ_3. Then, similarly to the proof of Theorem 3.2, the convergence (8.1) follows from (8.2), (5.3), $[A2]$ (ii) and consistency of the initial estimators. Remark that in the last part, we used tightness of $\big(\mathcal{Y}_n^{(3)}\big)_{n\in\mathbb{N}}$ derived by the estimate of $\partial_3 H$. □

To go further, some consideration is needed for the initial estimator $\hat{\theta}_2^0$. In our question, mathematically, the "martingale part" of $\hat{\theta}_2^0$ must be orthogonal to that of $\Delta_j Y$. Otherwise, the information of $\Delta_j X$ is mixed and it is out of the question. To avoid self-contradiction, we can not consider the initial estimator $\hat{\theta}_2^0$ that uses the first chaos of $(w_t)_{t\in[0,nh]}$, that gives additional information to $\Delta_j Y$. In order to understand this, it is necessary to go back to the proof of Theorem 3.6 though omitted in this article. In reality, the orthogonality means that the estimator $\hat{\theta}_2^0$ is a function of a prior information, e.g., an estimator based on the data sampled at time jh $(j=0,-1,-2,...,-m_n)$ when the process Z is extended as a stationary process on $\mathbb{R}$. The ideal case is we know the value θ_2^* a priori, that is, $\hat{\theta}_2^0=\theta_2^*$. Even when θ_2^* is unknown, if $\hat{\theta}_2^0$ is naturally constructed from the data Z_{t_j} $(j=0,-1,-2,...,-m_n)$ for example, the case where $m_n/n\to\infty$ is the same in the first order inference, while they are distinguishable in the higher-order inference (Sakamoto and Yoshida [41]). Then, under a certain set of conditions, we have

$$n^{1/2}h^{-1/2}\big(\hat{\vartheta}_3-\theta_3^*\big)\to^d N(0,4\Gamma_{33}^{-1}). \tag{8.4}$$

Therefore $\hat{\theta}_3^0$ is superior to $\hat{\vartheta}_3$. Remark that $\hat{\theta}_3^0$ was given an initial estimator of θ_2 with error rate $O_p((nh)^{-1/2})$ but not given the above $\hat{\theta}_2^0$ having the faster rate $O_p((m_nh)^{-1/2})$. In what follows, we will consider slightly more general $\hat{\theta}_2^0$ and show a convergence including (8.4) as a special case.

[A4$^\times$] **(i)** $\hat{\theta}_1^0-\theta_1^* \;=\; O_p(n^{-1/2})$ as $n\to\infty$.

(ii) $\hat{\theta}_2$ is $\mathcal{F}_0$-measurable and

$$\sqrt{nh}\big(\hat{\theta}_2-\theta_2^*\big)\to^d \mathbb{L}$$

as $n\to\infty$ for some random variable $\mathbb{L}$.

Lemma 8.2. *Suppose that* [A1] *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,2,2,1,3,1)$, [A2] *and* [A4$^\times$] *are satisfied. Then*

$$n^{-1/2}h^{1/2}\,\partial_3\mathcal{H}_n^{(3)}(\theta_3^*)\to^d \mathcal{N}+\mathcal{A}\mathbb{L} \tag{8.5}$$

as $n\to\infty$, *where* $\mathcal{N}\sim N(0,4^{-1}\Gamma_{33})$ *independent of* $\mathbb{L}$, *and*

$$\mathcal{A}=-\frac{3}{2}\int V(z,\theta_1^*,\theta_3^*)^{-1}\big[H_x(z,\theta_3^*)\partial_2A(z,\theta_2^*),\partial_3H(z,\theta_3^*)\big]\nu(dz).$$

Proof. We have the following decomposition:

$$\begin{aligned} n^{-1/2}h^{1/2}\,\partial_3\mathcal{H}_n^{(3)}(\theta_3^*) \;=\;& R_n^{(8.7)}(\hat{\theta}_1^0,\hat{\theta}_1^0)+R_n^{(8.8)}(\hat{\theta}_1^0,\hat{\theta}_2^0)\\ &+R_n^{(8.9)}(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_1^0,\hat{\theta}_2^0)+R_n^{(8.10)}(\hat{\theta}_1^0,\hat{\theta}_1^0,\hat{\theta}_1^0,\hat{\theta}_2^0) \end{aligned} \tag{8.6}$$

where

$$\begin{aligned} &R_n^{(8.7)}(\theta_1,\theta_1')\\ =\;& n^{-1/2}\sum_{j=1}^n 3V(Z_{t_{j-1}},\theta_1,\theta_3^*)^{-1}\Big[h^{-3/2}\big(\Delta_jY-hG_n(Z_{t_{j-1}},\theta_1',\theta_2^*,\theta_3^*)\big)\\ &\qquad\otimes\partial_3H(Z_{t_{j-1}},\theta_3^*)\big)\Big], \end{aligned} \tag{8.7}$$

$$\begin{aligned} R_n^{(8.8)}(\theta_1,\theta_2) \;=\;& -n^{-1/2}\sum_{j=1}^n 3V(Z_{t_{j-1}},\theta_1,\theta_3^*)^{-1}\\ &\cdot\Big[2^{-1}h^{1/2}H_x(Z_{t_{j-1}},\theta_3^*)\big[A(Z_{t_{j-1}},\theta_2)-A(Z_{t_{j-1}},\theta_2^*)\big]\\ &\qquad\otimes\partial_3H(Z_{t_{j-1}},\theta_3^*)\big)\Big], \end{aligned} \tag{8.8}$$

$$\begin{aligned}
&R_n^{(8.9)}(\theta_1,\theta_2,\theta_1',\theta_2')\\
=\ & n^{-1/2}\sum_{j=1}^n 3V(Z_{t_{j-1}},\theta_1,\theta_3^*)^{-1}\Big[h^{-3/2}\big(\Delta_j Y-hG_n(Z_{t_{j-1}},\theta_1',\theta_2',\theta_3^*)\big),\\
&\qquad 2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\theta_1,\theta_2,\theta_3^*)\big)\Big]
\end{aligned} \tag{8.9}$$

and

$$\begin{aligned}
&R_n^{(8.10)}(\theta_1,\theta_1',\theta_2')\\
=\ & \frac{3}{2}n^{-1/2}h^{1/2}\sum_{j=1}^n \big(V^{-1}(\partial_3 V)V^{-1}\big)(Z_{t_{j-1}},\theta_1,\theta_3^*)\\
&\cdot\big[\big\{h^{-3/2}\big(\Delta_j Y-hG_n(Z_{t_{j-1}},\theta_1',\theta_2',\theta_3^*)\big)\big\}^{\otimes 2}\\
&-3^{-1}V(Z_{t_{j-1}},\theta_1',\theta_3^*)\big].
\end{aligned} \tag{8.10}$$

Then we have the following estimates:

$$\begin{aligned}
&R_n^{(8.7)}(\hat\theta_1^0,\hat\theta_1^0)\\
=\ & R_n^{(8.7)}(\hat\theta_1^0,\theta_1^*)+O_p(h^{1/2})\quad \text{(Lemma 5.5 (b) and [A4}^{\times}\text{] (}i\text{))}\\
=\ & n^{-1/2}\sum_{j=1}^n 3V(Z_{t_{j-1}},\hat\theta_1^0,\theta_3^*)^{-1}\Big[h^{-3/2}\big(\xi_j^{(5.17)}+\xi_j^{(5.18)}\big),\\
&\qquad \partial_3 H(Z_{t_{j-1}},\theta_3^*)\big)\Big]+O_p(\sqrt{n}h)\\
&\qquad\qquad \text{(Lemma 5.5 (a))}\\
=\ & n^{-1/2}\sum_{j=1}^n 3V(Z_{t_{j-1}},\theta_1^*,\theta_3^*)^{-1}\\
&\cdot\Big[h^{-3/2}\kappa(Z_{t_{j-1}},\theta_1^*,\theta_3^*)\zeta_j,\ \partial_3 H(Z_{t_{j-1}},\theta_3^*)\big)\Big]\\
&+o_p(1)\quad \text{(random field argument, and orthogonality between } \{\xi_j^{(5.18)}\}_j\text{)}.
\end{aligned} \tag{8.11}$$

and

$$\begin{aligned}
&R_n^{(8.8)}(\hat\theta_1^0,\hat\theta_2^0)\\
=\ & -n^{-1/2}\sum_{j=1}^n 3V(Z_{t_{j-1}},\hat\theta_1^0,\theta_3^*)^{-1}\\
&\cdot\Big[2^{-1}h^{1/2}H_x(Z_{t_{j-1}},\theta_3^*)\big[A(Z_{t_{j-1}},\hat\theta_2^0)-A(Z_{t_{j-1}},\theta_2^*)\big],\ \partial_3 H(Z_{t_{j-1}},\theta_3^*)\big)\Big]\\
=\ & -n^{-1}\sum_{j=1}^n 3V(Z_{t_{j-1}},\hat\theta_1^0,\theta_3^*)^{-1}
\end{aligned}$$

$$\cdot\Big[2^{-1}H_x(Z_{t_{j-1}},\theta_3^*)\Big[\int_0^1 \partial_2 A(Z_{t_{j-1}},\theta_2^*+s(\hat{\theta}_2^0-\theta_2^*))ds\sqrt{nh}(\hat{\theta}_2^0-\theta_2^*)\Big],$$
$$\partial_3 H(Z_{t_{j-1}},\theta_3^*)\Big)\Big]$$
$$= \mathcal{A}\big[\sqrt{nh}(\hat{\theta}_2^0-\theta_2^*)\big]+o_p(1) \quad (\text{use } [A2] \text{ and } \partial_2^2 A). \tag{8.12}$$

Moreover,

$$\begin{aligned} R_n^{(8.9)}(\hat{\theta}_1^0,\hat{\theta}_2^0,\hat{\theta}_1^0,\hat{\theta}_2^0) &= R_n^{(8.9)}(\theta_1^*,\theta_2^*,\theta_1^*,\theta_2^*)+O_p(h) \\ &\qquad ([A4^\times], \text{ Lemmas 5.3 (b) and 5.5 (b)}) \\ &= O_p(h) \quad (\text{orthogonality and Lemma 5.3 (b)}) \end{aligned} \tag{8.13}$$

and

$$R_n^{(8.10)}(\hat{\theta}_1^0,\hat{\theta}_1^0,\hat{\theta}_2^0)=O_p(h^{1/2}). \tag{8.14}$$

We apply the martingale central limit theorem to (8.11), and use $[A4^\times]$ (ii) to (8.12). Then we obtain the convergence (8.5) from $[A4^\times]$, (8.6) and (8.11)-(8.14). □

Lemma 8.3. *Suppose that* $[A1]$ *with* $(i_A,j_A,i_B,j_B,i_H,j_H)=(1,1,2,1,3,2)$, $[A2]$ *and* $[A4^\times]$ *are satisfied. Then*

$$\sup_{\theta_3\in B_n}\Big|n^{-1}h\,\partial_3^2\mathcal{H}_n^{(3)}(\theta_3)+4^{-1}\Gamma_{33}\Big|\to^p 0$$

for any sequence of balls B_n *in* $\mathbb{R}^{\mathsf{p}_3}$ *shrinking to* θ_3^*.

Proof. A useful decomposition in this situation is:

$$\begin{aligned} n^{-1}h\,\partial_3^2\mathcal{H}_n^{(3)}(\theta_3) &= n^{-1}R_{33,1}(\theta_3)+n^{-1}h^{1/2}R_{33,2}(\theta_3)+n^{-1}hR_{33,3}(\theta_3) \\ &\quad +n^{-1}h^{1/2}R_{33,4}(\theta_3)+n^{-1}hR_{33,5}(\theta_3), \end{aligned}$$

where

$$\begin{aligned} R_{33,1}(\theta_3) &= -\sum_{j=1}^n 3V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)^{-1} \\ &\qquad \cdot\big[\big(\partial_3 H(Z_{t_{j-1}},\theta_3)+2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)\big)^{\otimes 2}\big], \end{aligned}$$

$$\begin{aligned} R_{33,2}(\theta_3) &= \sum_{j=1}^n 3V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)^{-1}\big[h^{-3/2}\big(\Delta_j Y-hG_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)\big), \\ &\qquad \partial_3^2 H(Z_{t_{j-1}},\theta_3)+2^{-1}h\partial_3^2 L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)\big], \end{aligned}$$

$$R_{33,3}(\theta_3)=-\frac{1}{2}\sum_{j=1}^n\big(V^{-1}(\partial_3 V)V^{-1}\big)(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)\big[\partial_3 V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)\big],$$

$$R_{33,4}(\theta_3) = -6\sum_{j=1}^{n}\big(V^{-1}(\partial_3 V)V^{-1}\big)(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3) \cdot\big[h^{-3/2}\big(\Delta_j Y - hG_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)\big), \partial_3 H(Z_{t_{j-1}},\theta_3) + 2^{-1}h\partial_3 L_H(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)\big]$$

and

$$R_{33,5}(\theta_3) = \frac{1}{2}\sum_{j=1}^{n}\partial_3\big\{\big(V^{-1}(\partial_3 V)V^{-1}\big)(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)\big\} \cdot\Big[3\big\{h^{-3/2}\big(\Delta_j Y - hG_n(Z_{t_{j-1}},\hat{\theta}_1^0,\hat{\theta}_2^0,\theta_3)\big)\big\}^{\otimes 2} - V(Z_{t_{j-1}},\hat{\theta}_1^0,\theta_3)\Big].$$

Obviously,

$$\sup_{\theta_3\in\overline{\Theta}_3}\big|n^{-1}hR_{33,3}(\theta_3)\big| = O_p(h)$$

By using Lemma 5.5 (b) with [$A4^{\times}$] (only rates of convergence), Lemma 5.3 (b), and orthogonality, we see

$$\sup_{\theta_3\in B_n}\big|n^{-1}h^{1/2}R_{33,4}(\theta_3)\big| = O_p(n^{-1/2}h^{1/2}) + O_p(\mathrm{diam}B_n) + O_p(h) \ = \ o_p(1)$$

and similarly

$$\sup_{\theta_3\in B_n}\big|n^{-1}hR_{33,5}(\theta_3)\big| = O_p(n^{-1/2}h) + O_p(\mathrm{diam}B_n) + O_p(h) \ = \ o_p(1).$$

Estimate of $R_{33,2}(\theta_3)$ is similar to that of $R_{33,4}(\theta)$:

$$\sup_{\theta_3\in B_n}\big|n^{-1}h^{1/2}R_{33,2}(\theta_3)\big| = o_p(1).$$

Using $\partial_3^i H$ $(i=1,2)$ and $\partial_3 H_x$ for tightness, we obtain

$$\sup_{\theta_3\in B_n}\big|n^{-1}R_{33,1}(\theta_3) - n^{-1}R_{33,1}(\theta_3^*)\big| = o_p(1).$$

Moreover, it is easy to see

$$\sup_{\theta_3\in B_n}\big|n^{-1}R_{33,1}(\theta_3^*) + 4^{-1}\Gamma_{33}\big| = o_p(1).$$

Thus, the proof is completed. □

Remark that Condition [$A4$] is sufficient for Lemma 8.3 since only rate of convergence of $\hat{\theta}_1^0$ and $\hat{\theta}_2^0$ used in the above proof. Denote by $\mathcal{L}\{\xi\}$ the distribution of a random variable ξ. A "convolution theorem" follows from Proposition 8.1 and Lemmas 8.2 and 8.3.

Theorem 8.4. *Suppose that* [*A*1] *with* $(i_A, j_A, i_B, j_B, i_H, j_H) = (1, 2, 2, 1, 3, 2)$, [*A*2], [*A*3] (*iii*) *and* [*A*4$^{\times}$] *are satisfied. Then*

$$n^{1/2}h^{-1/2}(\hat{\vartheta}_3 - \theta_3^*) \to^d N(0, 4\Gamma_{33}^{-1}) * \mathcal{L}\{4\Gamma_{33}^{-1}\mathcal{A}\mathbb{L}\}$$

as $n \to \infty$.

The convergence (8.4) is a special case of Theorem 8.4 when $\mathbb{L} = 0$ a.s. Like Hájek, the result shows the superiority of $\hat{\theta}_3^0$ to $\hat{\vartheta}_3$.

Apart from this problem, we can also ask what occurs if we use $\hat{\theta}_2^0$ depending on $(X_{t_j})_{j=0,\ldots,n}$. It is natural but another question. The resulting $\hat{\vartheta}_3$ obviously exploits $\Delta_j X$ through $\hat{\theta}_2^0$. Moreover, $\hat{\vartheta}_3$ must initially pay $4\Gamma_{33}^{-1}$, as it can be observed in the above proof. This means $\hat{\vartheta}_3$ does not dominate $\hat{\theta}_3^0$, besides, if $\hat{\theta}_2^0$ performs well having a small asymptotic variance, then $\hat{\vartheta}_3$ is not admissible (in the sense of the decision theory) at least locally in Θ_3. Furthermore, the behavior of $\hat{\vartheta}_3$ strongly depends on that of the error of $\hat{\theta}_2^0$. In other words, the asymptotic property of $\hat{\vartheta}_3$ cannot separate from the properties of the initial estimators even when their convergence rates are ensured.

References

[1] Adams, R. A. (1975). *Sobolev spaces.* Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London Pure and Applied Mathematics, Vol. 65. MR0450957 (56 ##9247)

[2] Bibby, B. M. and Sørensen, M. (1995). Martingale estimation functions for discretely observed diffusion processes. *Bernoulli* **1** 17–39. MR1354454 (97c:62201)

[3] Comte, F. and Genon-Catalot, V. (2006). Penalized projection estimator for volatility density. *Scand. J. Statist.* **33** 875–893. MR2300921

[4] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli* **13** 514–543. MR2331262

[5] Comte, F., Prieur, C. and Samson, A. (2017). Adaptive estimation for stochastic damping Hamiltonian systems under partial observation. *Stochastic processes and their applications* **127** 3689–3718. MR3707242

[6] Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *Ann. Statist.* **33** 2507–2528. MR2253093

[7] Dalalyan, A. and Reiss, M. (2006). Asymptotic statistical equivalence for scalar ergodic diffusions. *Probab. Theory Related Fields* **134** 248–282. MR2222384

[8] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Related Fields* **137** 25–47. MR2278451

[9] Dalalyan, A. S. and Kutoyants, Y. A. (2002). Asymptotically efficient trend coefficient estimation for ergodic diffusion. *Math. Methods Statist.* **11** 402–427 (2003). MR1979742

[10] De Gregorio, A. and Iacus, S. M. (2012). Adaptive LASSO-type estimation for multivariate diffusion processes. *Econometric Theory* **28** 838–860. MR2959127

[11] Delattre, S., Gloter, A. and Yoshida, N. (2020). Rate of Estimation for the Stationary Distribution of Stochastic Damping Hamiltonian Systems with Continuous Observations. *arXiv preprint arXiv:2001.10423*.

[12] Ditlevsen, S. and Samson, A. (2019). Hypoelliptic diffusions: filtering and inference from complete and partial observations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **81** 361–384. MR3928146

[13] Eguchi, S., Masuda, H. et al. (2018). Schwarz type model comparison for LAQ models. *Bernoulli* **24** 2278–2327. MR3757530

[14] Genon-Catalot, V., Jeantheau, T. and Laredo, C. (1999). Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5** 855–872. MR1715442

[15] Genon-Catalot, V. and Larédo, C. (2016). Estimation for stochastic differential equations with mixed effects. *Statistics* **50** 1014–1035. MR3522386

[16] Gloter, A. (2000). Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM: Probability and Statistics* **4** 205–227. MR1808927

[17] Gloter, A. (2001). Parameter estimation for a discrete sampling of an intergrated Ornstein-Uhlenbeck process. *Statistics* **35** 225–243. MR1925514

[18] Gloter, A. (2006). Parameter estimation for a discretely observed integrated diffusion process. *Scandinavian Journal of Statistics* **33** 83–104. MR2255111

[19] Gloter, A. (2007). Efficient estimation of drift parameters in stochastic volatility models. *Finance and Stochastics* **11** 495–519. MR2335831

[20] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. I. Local asymptotic normality. *ESAIM: Probability and Statistics* **5** 225–242. MR1875672

[21] Gloter, A. and Jacod, J. (2001). Diffusions with measurement errors. II. Optimal estimators. *ESAIM: Probability and Statistics* **5** 243–260. MR1875673

[22] Gloter, A. and Yoshida, N. (2020). Adaptive and non-adaptive estimation for degenerate diffusion processes. *arXiv preprint arXiv:2002.10164*.

[23] Kamatani, K. and Uchida, M. (2014). Hybrid multi-step estimators for stochastic differential equations based on sampled data. *Statistical Inference for Stochastic Processes* **18** 177–204. MR3348584

[24] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24** 211–229. MR1455868 (99c:62234)

[25] Küchler, U. and Sorensen, M. (1997). *Exponential families of stochastic processes* **3**. Springer. MR1458891

[26] Kusuoka, S. and Yoshida, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probab. Theory Related Fields* **116** 457–484. MR1757596 (2002a:60093)

[27] Kutoyants, Y. A. (1984). *Parameter estimation for stochastic processes*. *Research and Exposition in Mathematics* **6**. Heldermann Verlag, Berlin. Translated from the Russian and edited by B. L. S. Prakasa Rao. MR777685 (86b:62145)

[28] Kutoyants, Y. A. (1997). Some problems of nonparametric estimation by observations of ergodic diffusion process. *Statist. Probab. Lett.* **32** 311–320. MR1440841 (98i:62031)

[29] Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. *Springer Series in Statistics*. Springer-Verlag London Ltd., London. MR2144185 (2006b:62005)

[30] León, J. R. and Samson, A. (2018). Hypoelliptic stochastic FitzHugh–Nagumo neuronal model: Mixing, up-crossing and estimation of the spike rate. *The Annals of Applied Probability* **28** 2243–2274. MR3843828

[31] Masuda, H. (2013). Convergence of Gaussian quasi-likelihood random fields for ergodic Lévy driven SDE observed at high frequency. *The Annals of Statistics* **41** 1593–1641. MR3113823

[32] Masuda, H. and Shimizu, Y. (2017). Moment convergence in regularized estimation under multiple and mixed-rates asymptotics. *Mathematical Methods of Statistics* **26** 81–110. MR3667404

[33] Melnykova, A. (2019). Parametric inference for multidimensional hypoelliptic ergodic diffusion with full observations. working paper or preprint. MR4136706

[34] Meyn, S. P. and Tweedie, R. L. (1993). Stability of Markovian processes. III. Foster-Lyapunov criteria for continuous-time processes. *Adv. in Appl. Probab.* **25** 518–548. MR1234295 (94g:60137)

[35] Nakakita, S. H., Kaino, Y. and Uchida, M. (2020). Quasi-likelihood analysis and Bayes-type estimators of an ergodic diffusion plus noise. *Annals of the Institute of Statistical Mathematics* 1–49. MR4205246

[36] Nakakita, S. H. and Uchida, M. (2019). Inference for ergodic diffusions plus noise. *Scandinavian Journal of Statistics* **46** 470–516. MR3948565

[37] Nualart, D., Yoshida, N. et al. (2019). Asymptotic expansion of Skorohod integrals. *Electronic Journal of Probability* **24**. MR4029422

[38] Pardoux, E. and Veretennikov, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *Annals of probability* 1061–1085. MR1872736

[39] Prakasa Rao, B. L. S. (1983). Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Math. Operationsforsch. Statist. Ser. Statist.* **14** 195–209. MR704787 (84k:62122)

[40] Prakasa Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. In *Statistical inference from stochastic processes (Ithaca, NY, 1987)*. *Contemp. Math.* **80** 249–284. Amer. Math. Soc., Providence, RI. MR999016 (90h:62197)

[41] Sakamoto, Y. and Yoshida, N. (2009). Third-order asymptotic expansion of M-estimators for diffusion processes. *Ann. Inst. Statist. Math.* **61** 629–661. MR2529969

[42] Samson, A. and Thieullen, M. (2012). A contrast estimator for com-

pletely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications* **122** 2521–2552. MR2926166

[43] Schmisser, E. (2013). Penalized nonparametric drift estimation for a multidimensional diffusion process. *Statistics* **47** 61–84. MR3023015

[44] Suzuki, T. and Yoshida, N. (2020). Penalized least squares approximation methods and their applications to stochastic processes. *Japanese Journal of Statistics and Data Science* 1–29. MR4181988

[45] Uchida, M. (2010). Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Annals of the Institute of Statistical Mathematics* **62** 161–187. MR2577445

[46] Uchida, M. and Yoshida, N. (2001). Information criteria in model selection for mixing processes. *Stat. Inference Stoch. Process.* **4** 73–98. MR1850590 (2002g:62005)

[47] Uchida, M. and Yoshida, N. (2011). Estimation for misspecified ergodic diffusion processes from discrete observations. *ESAIM: Probability and Statistics* **15** 270–290. MR2870516

[48] Uchida, M. and Yoshida, N. (2012). Adaptive estimation of an ergodic diffusion process based on sampled data. *Stochastic Processes and their Applications* **122** 2885–2924. MR2931346

[49] Wu, L. (2001). Large and moderate deviations and exponential convergence for stochastic damping Hamiltonian systems. *Stochastic processes and their applications* **91** 205–238. MR1807683

[50] Yoshida, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.* **42** 221–251. MR1064786 (92a:62130)

[51] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.* **41** 220–242. MR1172898 (93g:62113)

[52] Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Annals of the Institute of Statistical Mathematics* **63** 431–479. MR2786943

[53] Yoshida, N. (2013). Martingale expansion in mixed normal limit. *Stochastic Processes and their Applications* **123** 887–933. MR3005009