Stochastic Processes and their Applications 174 (2024) 104384

Contents lists available at ScienceDirect

Stochastic Processes and their Applications

journal homepage: www.elsevier.com/locate/spa

Parameter inference for degenerate diffusion processes

Yuga Iguchi [a,*], Alexandros Beskos [a], Matthew M. Graham [b]

[a] Department of Statistical Science, University College London, 1-19 Torrington Place, WC1E 6BT, London, United Kingdom
[b] Advanced Research Computing Centre, University College London, Bidborough House, WC1H 9BT, London, United Kingdom

ARTICLE INFO

Keywords:
Stochastic differential equation
Hypo-elliptic diffusion
Hörmander's condition
Partial observations
Generalised Langevin equation

ABSTRACT

We study parametric inference for ergodic diffusion processes with a degenerate diffusion matrix. Existing research focuses on a particular class of hypo-elliptic Stochastic Differential Equations (SDEs), with components split into 'rough'/'smooth' and noise from rough components propagating directly onto smooth ones, but some critical model classes arising in applications have yet to be explored. We aim to cover this gap, thus analyse the *highly degenerate* class of SDEs, where components split into further sub-groups. Such models include e.g. the notable case of generalised Langevin equations. We propose a tailored time-discretisation scheme and provide asymptotic results supporting our scheme in the context of high-frequency, full observations. The proposed discretisation scheme is applicable in much more general data regimes and is shown to overcome biases via simulation studies also in the practical case when only a smooth component is observed. Joint consideration of our study for highly degenerate SDEs and existing research provides a general 'recipe' for the development of time-discretisation schemes to be used within statistical methods for general classes of hypo-elliptic SDEs.

1. Introduction

This work addresses the statistical calibration of a wide class of hypo-elliptic diffusions. Stochastic Differential Equations (SDEs) are widely used as an effective tool to describe dynamics of the time evolution of phenomena of interest across a multitude of disciplines. Consider SDE models of the following general form:

$$dX_t = V_0(X_t, \theta)dt + \sum_{j=1}^{d} V_j(X_t, \theta)dB_{j,t}, \qquad X_0 = x \in \mathbb{R}^N, \tag{1}$$

with $V_j(\cdot,\theta) : \mathbb{R}^N \to \mathbb{R}^N$, $0 \le j \le d$, for parameter θ, driven by the d-dimensional standard Brownian motion $B = (B_{1,t}, \dots, B_{d,t})$, $t \ge 0$, defined upon the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\ge 0}, \mathbb{P})$, with $d, N \ge 1$. Several theoretical results about parameter inference for SDEs have been established under positive definiteness conditions on the diffusion matrix $a = VV^{\top} \in \mathbb{R}^{N\times N}$, with $V = [V_1, \dots, V_d]$. In such a case, the solution of (1) is referred to as an *elliptic* diffusion. However, many important applications give rise to diffusion processes that allow matrix a to be degenerate. We give below examples for such classes of SDEs. Under the weak Hörmander's condition, discussed later in this work, the process defined via the SDE (1) with degenerate diffusion matrix a permits a density with respect to (w.r.t.) the Lebesgue measure for its transition dynamics, and is referred to as a *hypo-elliptic* diffusion.

* Corresponding author.
E-mail address: yuga.iguchi.21@ucl.ac.uk (Y. Iguchi).

https://doi.org/10.1016/j.spa.2024.104384
Received 23 August 2023; Received in revised form 28 February 2024; Accepted 13 May 2024
Available online 17 May 2024

1.1. Classes of diffusion models

We can summarise the SDE models we consider in this work via two classes of hypo-elliptic diffusions. The first hypo-elliptic class is determined via the following degenerate SDE:

$$dX_t = \begin{bmatrix} dX_{S,t} \\ dX_{R,t} \end{bmatrix} = \begin{bmatrix} V_{S,0}(X_t,\beta_S) \\ V_{R,0}(X_t,\beta_R) \end{bmatrix} dt + \sum_{j=1}^{d} \begin{bmatrix} \mathbf{0}_{N_S} \\ V_{R,j}(X_t,\sigma) \end{bmatrix} dB_{j,t}, \qquad X_0 = x = \left[x_S^\top, x_R^\top\right]^\top \in \mathbb{R}^N. \tag{Hypo-I}$$

Here, the involved SDE functionals are specified as follows:

$$V_{S,0} : \mathbb{R}^N \times \Theta_{\beta_S} \to \mathbb{R}^{N_S}, \qquad V_{R,0} : \mathbb{R}^N \times \Theta_{\beta_R} \to \mathbb{R}^{N_R}, \qquad V_{R,j} : \mathbb{R}^N \times \Theta_\sigma \to \mathbb{R}^{N_R}, \quad 1 \le j \le d,$$

for positive integers N_S, N_R such that $N = N_S + N_R$, and unknown parameter vector

$$\theta = (\beta_S, \beta_R, \sigma) \in \Theta = \Theta_{\beta_S} \times \Theta_{\beta_R} \times \Theta_\sigma \subseteq \mathbb{R}^{N_{\beta_S}} \times \mathbb{R}^{N_{\beta_R}} \times \mathbb{R}^{N_\sigma},$$

for positive integers N_{β_S}, N_{β_R}, N_σ, and a compact set Θ. For class (Hypo-I), we will later on introduce a condition upon the vector-valued functionals $\{V_{S,0}, V_{R,0}, \dots, V_{R,d}\}$ that is sufficient for the law of X_t to admit a Lebesgue density, and which is related to the weak Hörmander's condition. In brief, the condition stipulates that $X_{R,t}$ is indeed a rough component, and that *all* coordinates of the drift function $V_{S,0}(X_t, \beta_S)$ properly relate with the rough component $X_{R,t}$, so that randomness from $X_{R,t}$ is propagated onto all coordinates of vector $X_{S,t}$.

The development of a theoretical and algorithmic framework for parametric inference over class (Hypo-I) has been the topic of several recent works, see e.g. [8,15,18,32]. However, we stress that several important hypo-elliptic SDEs used in practice do not belong in class (Hypo-I) – in the sense that *not all* coordinates of the drift function for the smooth components involve the rough component $X_{R,t}$ – thus are not covered by recent investigations. We specify a key class of practically useful but under-explored hypo-elliptic SDEs via the following equation:

$$\begin{aligned} dX_t &= \begin{bmatrix} dX_{S_1,t} \\ dX_{S_2,t} \\ dX_{R,t} \end{bmatrix} = \begin{bmatrix} V_{S_1,0}(X_{S_1,t}, X_{S_2,t}, \beta_{S_1}) \\ V_{S_2,0}(X_t, \beta_{S_2}) \\ V_{R,0}(X_t, \beta_R) \end{bmatrix} dt + \sum_{j=1}^{d} \begin{bmatrix} \mathbf{0}_{N_{S_1}} \\ \mathbf{0}_{N_{S_2}} \\ V_{R,j}(X_t,\sigma) \end{bmatrix} dB_{j,t}; \\ X_0 &= x = \left[x_{S_1}^\top, x_{S_2}^\top, x_R^\top\right]^\top \in \mathbb{R}^N. \end{aligned} \tag{Hypo-II}$$

Thus, the drift functions of the smooth components are now specified as:

$$V_{S_1,0} : \mathbb{R}^{N_{S_1}+N_{S_2}} \times \Theta_{\beta_{S_1}} \to \mathbb{R}^{N_{S_1}}, \qquad V_{S_2,0} : \mathbb{R}^N \times \Theta_{\beta_{S_2}} \to \mathbb{R}^{N_{S_2}},$$

and the parameter vector writes as:

$$\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta = \Theta_{\beta_{S_1}} \times \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma \subseteq \mathbb{R}^{N_{\beta_{S_1}}} \times \mathbb{R}^{N_{\beta_{S_2}}} \times \mathbb{R}^{N_{\beta_R}} \times \mathbb{R}^{N_\sigma},$$

for positive integers N_{S_1}, N_{S_2}, $N_{\beta_{S_1}}$, $N_{\beta_{S_2}}$, and $N = N_{S_1} + N_{S_2} + N_R$, where Θ is again a compact set. We stress again that the drift function $V_{S_1,0}(X_{S_1,t}, X_{S_2,t}, \beta_{S_1})$ does not depend on the rough component $X_{R,t}$, so noise from $X_{R,t}$ is not directly propagated onto $X_{S_1,t}$. Thus, (Hypo-II) is treated as a different model class than (Hypo-I). We refer to (Hypo-II) as the *highly degenerate* class of SDEs.

For model class (Hypo-II) we will later on set-up a restriction over its constituent vector-valued functionals $\{V_{S_1,0}, V_{S_2,0}, V_{R,0}, \dots, V_{R,d}\}$, related to Hörmander's condition, which will differ from the corresponding condition assumed for (Hypo-I). Roughly, such a requirement will guarantee that noise from the rough component $X_{R,t}$ indeed propagates onto all coordinates of $X_{S_2,t}$, first, then moving onto $X_{S_1,t}$. Thus, $X_{S_1,t}$ is 'smoother' than $X_{S_2,t}$. Importantly, a consequence of such a behaviour is that class (Hypo-II) is *not* included within (Hypo-I), instead the two classes, (Hypo-I) and (Hypo-II), are intrinsically distinct and must be treated separately in terms of theoretical and algorithmic considerations.

1.2. A motivating class of models

The *non-Markovian Langevin equation* (or *generalised Langevin equation* (GLE)) is used in a wide range of applications due to its effectiveness in describing complex stochastic systems with memory effects (thus, of non-Markovian structure). Examples include dynamics observed in protein folding [1], cancer cells [26], flocks of birds [12], molecules [27] and coarse-grained systems [20,25]. For simplicity, we consider here a one-dimensional particle with unit mass, and denote its position and momentum, respectively, by (q,p). Then, a GLE describes the particle dynamics as follows:

$$\begin{aligned} \dot{q}_t &= p_t; \\ \dot{p}_t &= -U'(q_t) - \int_0^t K(t-s) p_s ds + \eta_t, \end{aligned} \tag{GLE}$$

where $U : \mathbb{R} \to \mathbb{R}$ is an appropriate potential function, $K : [0,\infty) \to \mathbb{R}$ is the *memory kernel*, and η_t is a zero-mean stationary Gaussian noise with auto-correlation specified via a fluctuation–dissipation relation in equilibrium, i.e., $\mathbb{E}[\eta_t \eta_s] = K(t-s)$, $s,t > 0$, given a unit temperature. Due to the presence of $K(\cdot)$, particle dynamics will depend on the full state history, with such a property being quite desirable in applications, see e.g. the references given above. However, the cost of generating the dynamics of model (GLE) can be overly expensive. Thus, a standard approach followed in practice is to introduce a parametrisation for the memory kernel $K(\cdot)$ and represent the non-Markovian system (GLE) as a Markovian one on an extended space, the latter system being referred to as *Quasi-Markovian Generalised Langevin Equation (QGLE)*. Such parametrisation is extremely rich, thus being able to accurately capture the behaviour of systems with general true kernel $K(\cdot)$. In particular, a common parametrisation of the memory kernel is the following:

$$K(t) = \alpha\delta(t) - \langle e^{-tA}\lambda, \lambda\rangle, \qquad \alpha > 0, \quad \lambda \in \mathbb{R}^m, \quad A \in \mathbb{R}^{m\times m}, \quad m \geq 1,$$

with $\delta = \delta(\cdot)$ the Dirac function. In this case, the original system in (GLE) can be equivalently re-written as the following Markovian one:

$$\begin{bmatrix} dq_t \\ dp_t \\ ds_t \end{bmatrix} = \begin{bmatrix} p_t \\ -U'(q_t) - \alpha p_t - \langle \lambda, s_t\rangle \\ -p_t\lambda - As_t \end{bmatrix} dt + \sum_{j=1}^{m+1} \begin{bmatrix} 0 \\ \sigma_j \end{bmatrix} dB_{j,t}, \qquad s_0 \sim \mathscr{N}(\mathbf{0}_m, I_m), \tag{QGLE-I}$$

with $s_t \in \mathbb{R}^m$ an auxiliary component and $\sigma_j \in \mathbb{R}_+^{m+1}$, $1 \leq j \leq m$. Another typical choice for the memory kernel is the following:

$$K(t) = \langle e^{-tA}\lambda, \lambda\rangle,$$

in which case the equivalent QGLE writes as:

$$\begin{bmatrix} dq_t \\ dp_t \\ ds_t \end{bmatrix} = \begin{bmatrix} p_t \\ -U'(q_t) + \langle \lambda, s_t\rangle \\ -p_t\lambda - As_t \end{bmatrix} dt + \sum_{j=1}^{m} \begin{bmatrix} 0 \\ 0 \\ \sigma_j \end{bmatrix} dB_{j,t}, \qquad s_0 \sim \mathscr{N}(\mathbf{0}_m, I_m), \tag{QGLE-II}$$

with $\sigma_j \in \mathbb{R}_+^m$. Class (QGLE-I) is investigated, e.g., in [5]. Then, class (QGLE-II) is popular, e.g., in thermodynamics modelling, see [23,29]. Class (QGLE-I) belongs in (Hypo-I), with the rough component comprised of p_t, s_t. For class (QGLE-II), the rough component consists only of s_t, with q_t depending on the smooth component p_t and not on s_t. Thus, (QGLE-II) lies within class (Hypo-II). Recently, parametric inference for GLEs within the QGLE setting, under discrete-time observations of the smooth component q_t, has been of interest for applications, see e.g. [12,37].

1.3. Related works and objectives

In this paper we investigate parameter estimation for the two classes of degenerate diffusion processes, (Hypo-I) and (Hypo-II), given discrete-time observations obtained at instances $0 \leq t_0 < t_1 < \cdots < t_n$, $n \in \mathbb{N}$, with equidistant observation intervals $\Delta_n := t_i - t_{i-1}$, $1 \leq i \leq n$. In particular, we consider the following scenarios for the observations:

1. *Complete observation regime*, i.e., with all N coordinates of X_t being observed.
2. *Partial observation regime*, i.e., with a strict subset of coordinates being observed. In agreement with applications, in this setting only the upper-most smooth component is assumed to be observed.

Within class (Hypo-I), and for the complete observation regime, [8,14,15] develop judicious discrete-time (conditionally) Gaussian approximations for the transition distribution. Such a proxy provides contrast estimators proven to be asymptotically normal in a high-frequency observation setting, i.e., $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, with a requirement that the step-size scales as $\Delta_n = o(n^{-1/2})$. It should be remarked that [8] considered hypo-elliptic SDEs in (Hypo-I) with particular restrictions, e.g., $N_S = 1$, and proposed contrast estimators separately for $\beta_S \in \Theta_{\beta_S}$ and $(\beta_R, \sigma) \in \Theta_{\beta_R}\times\Theta_\sigma$, while [14,15] provided a joint contrast estimator for all parameter $\theta = (\beta_S, \beta_R, \sigma)$ under a more general model structure. In [18], the step-size condition $\Delta_n = o(n^{-1/2})$ to obtain asymptotic normality is weakened to $\Delta_n = o(n^{-1/3})$. In the partial observation regime, with the upper-most smooth component being observed, the missing components must be carefully imputed given the available observations. For SDEs in class (Hypo-I), it is often the case that the dynamics of the smooth component is determined as

$$dX_{S,t} = X_{R,t}dt. \tag{2}$$

Such a remark also applies for class (Hypo-II), with the role of $X_{S,t}$ taken up by the upper-most smooth component, and the one of $X_{R,t}$ by the second smooth component. We keep the discussion within class (Hypo-I), as this is the context typically looked at in earlier literature. For the described setting, it is tempting and, indeed, widely used in practice, to recover the hidden rough component via finite-differences, using the observations $\{X_{S,t_{i+1}}\}_i$, i.e. via $X_{R,t_i} = (X_{S,t_{i+1}} - X_{S,t_i})/\Delta_n$, if the step-size Δ_n is small enough. However, [32,33] show that, in the context of bivariate models within class (Hypo-I), such an approach delivers (asymptotically) biased estimates of the diffusion parameter σ. To side-step the bias, [33] further developed a corrected contrast function built upon a conditionally Gaussian approximation for the rough component $X_{R,t}$. [8,32] argue against applying finite-differences and, instead, consider appropriate Itô–Taylor schemes leading to non-degenerate conditionally Gaussian approximations for the SDE transition density. Such proxies are then embedded within MCMC Gibbs samplers or Monte-Carlo Expectation-Maximisation (MC-EM) methods to impute the missing components conditionally on observations. Note that the approach does not require (2) for the smooth component. [32] then illustrate empirically that the scheme omitting drift terms of size $\mathcal{O}(\Delta_n^2)$ from the

Table 1
Parametric inference for hypo-elliptic SDEs from high-frequency, complete observations.

Work	Model	Joint estimation	Step-size condition
[8]	(Hypo-I) with $N_S = 1$ and diagonal diffusion matrix	×	$\Delta_n = o(n^{-1/2})$
[14,15]	(Hypo-I)	∘	$\Delta_n = o(n^{-1/2})$
[18]	(Hypo-I)	∘	$\Delta_n = o(n^{-1/3})$
This paper	(Hypo-II)	∘	$\Delta_n = o(n^{-1/2})$

Table 2
Parametric inference for hypo-elliptic SDEs from partial observations.

Work	Model	Approach	Estimation
[32]	(Hypo-I) satisfying (2) with $N_S = N_R = 1$	Finite-differences Gaussian approximation	Bias in σ Bias in β_R
[33]	(Hypo-I) satisfying (2) with $N_S = N_R = 1$	Finite-differences with a corrected contrast function	Unbiased
[8]	(Hypo-I) with $N_S = 1$ and diagonal diffusion matrix	Gaussian approximation	Unbiased
This paper	(Hypo-II)	Gaussian approximation	Unbiased

Itô–Taylor expansion of the smooth component leads to a biased estimation for the drift parameter, β_R, of the rough component. Subsequent analytical works [8,14,18] illustrated that the bias is resolved by adding the drift terms of size $\mathcal{O}(\Delta_n^2)$ in the smooth component within class (Hypo-I), as such terms are needed to counterbalance the noise terms of size $\mathcal{O}(\Delta_n^{3/2})$ arising in such an expansion.

The above discussion suggests that one of the possible recipes for accurate estimation of both hidden components and parameters would be the development of a conditionally Gaussian approximation for the full coordinates (as such Gaussianity allows for access to computationally effective inference methodologies) obtained via careful inclusion of higher-order terms from the relevant Itô–Taylor expansion. Such an insight for the design of a 'correct' discretisation scheme for the purposes of statistical inference has, arguably, not been clearly spelled out in the literature. Furthermore, the benefit of developing such a (locally) Gaussian discretisation is that it produces a closed-form approximate likelihood which is widely applicable to 'likelihood-based' inference of degenerate diffusions under both a frequentist/Bayesian framework. We hereby mention that other numerical schemes based on operator splitting (not on Itô–Taylor expansion) can be utilised in 'likelihood-free' inference approaches where only the simulation of SDE paths is required. Recently, exploiting the measure-preserving property of splitting schemes, several works [3,9] have presented empirical results showcasing that embedding such schemes in likelihood-free inference can lead to effective parameter estimation for a partially observed hypo-elliptic diffusion belonging in (Hypo-I). [31] developed an asymptotically unbiased contrast estimator based on splitting schemes for elliptic SDEs having a super-linear drift function and an additive noise, i.e., constant diffusion coefficients.

Our work aims to provide a comprehensive study of statistical calibration for a wide class of degenerate diffusion models. For this purpose, we review previous works for class (Hypo-I), and, then, we establish new analytical results for class (Hypo-II). The main contributions of our work can be summarised as follows:

(i) For the highly degenerate class (Hypo-II), we construct a conditionally Gaussian time-discretisation scheme. The corresponding transition density is well-defined (i.e. non-degenerate) under a suitable assumption on functionals $\{V_{S_{1},0}, V_{S_{2},0}, V_{R,0}, \ldots, V_{R,d}\}$ motivated both by modelling considerations and by adherence to the weak Hörmander's condition. We refer to the new proxy as the 'locally Gaussian scheme' in agreement with the name assigned by [15] to a conditionally Gaussian scheme developed for class (Hypo-I).

(ii) For class (Hypo-II), we define a joint contrast estimator based on the transition density of the locally Gaussian scheme. Then, we show that the estimator is asymptotically normal in the complete observation regime, under a high-frequency observation setting, i.e., for $n \to \infty$, $\Delta_n \to 0$, $n\Delta_n \to \infty$, with the additional condition that the step-size must scale as $\Delta_n = o(n^{-1/2})$.

(iii) Under the partial observation regime often encountered in practical applications, we show via analytical consideration of some case studies that use of a finite-difference method for estimation of hidden components leads to asymptotically biased estimation of the diffusion parameter σ for class (Hypo-II). Thus, we put forward the developed locally Gaussian scheme for (Hypo-II) as an effective tool to impute hidden components and estimate parameters.

(iv) By reviewing the methodology already produced in the literature for class (Hypo-I) and examining the new one produced in this work for (Hypo-II), we can provide a complete guideline for the development of a discretisation scheme for general degenerate diffusion processes so that the corresponding contrast function does not introduce bias in parameter estimation procedures.

A comparison of our work with early works is briefly summarised in Tables 1 and 2. The rest of the paper is organised as follows. Section 2 specifies the class of hypo-elliptic SDEs of relevance for this work, with reference to Hörmander's condition. Section 3 revisits the correct (in terms of its statistical properties) discretisation scheme for class (Hypo-I) and introduces the one for (Hypo-II). Section 4 provides our core analytical results of asymptotic consistency and normality for the statistical estimates obtained via the new scheme, in a complete observation setting. All proofs are collected in an Appendix A. We present case studies showcasing the emergence of bias when standard alternative schemes are called upon or when finite-differences are used to impute unobserved components (a common practice in applications). Under the correct schemes shown here for classes (Hypo-I) and (Hypo-II), we set up a simple Kalman filter for fitting a non-linear sub-class of models commonly arising in applications (we term these *conditional Gaussian non-linear systems*) in the practical partial observation setting. Section 5 presents numerical studies, for the partial observation regime, both for simple models and ones relevant to real applications, within class (Hypo-II). The code used in the numerical studies is available at https://github.com/YugaIgu/calibration-hypoSDEs. We finish with some conclusions in Section 6.

Notation. For the highly degenerate class (Hypo-II), to establish a common notation with (Hypo-I), we use the argument $x_S = (x_{S_1}, x_{S_2})$ and also set:

$$\begin{aligned} X_{S,t} &= \left[X_{S_1,t}^\top, X_{S_2,t}^\top\right]^\top \in \mathbb{R}^{N_S}, \quad N_S = N_{S_1} + N_{S_2}; \\ \beta_S &= \left[\beta_{S_1}^\top, \beta_{S_2}^\top\right]^\top \in \Theta_{\beta_S}, \quad \Theta_{\beta_S} = \Theta_{\beta_{S_1}} \times \Theta_{\beta_{S_2}}, \quad N_{\beta_S} = N_{\beta_{S_1}} + N_{\beta_{S_2}}; \\ V_{S,0}(x, \beta_S) &= \left[V_{S_1,0}(x_S, \beta_{S_1})^\top,\ V_{S_2,0}(x, \beta_{S_2})^\top\right]^\top. \end{aligned}$$

For $x \in \mathbb{R}^N$ and $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$, we write

$$\begin{aligned} V_0(x,\theta) &= \left[V_{S,0}(x,\beta_S)^\top,\ V_{R,0}(x,\beta_R)^\top\right]^\top; \\ V_j(x,\theta) &= \left[\mathbf{0}_{N_S}^\top,\ V_{R,j}(x,\sigma)^\top\right]^\top, \qquad 1 \le j \le d. \end{aligned} \tag{3}$$

For $\varphi(\cdot, \theta) : \mathbb{R}^N \to \mathbb{R}$, $\theta \in \Theta$, bounded up to 2nd order derivatives, we define the differential operators $\mathcal{L}$ and $\mathcal{L}_j$, $1 \le j \le d$, as:

$$\begin{aligned} \mathcal{L}\varphi(x,\theta) &= \sum_{i=1}^N V_0^i(x,\theta)\frac{\partial \varphi}{\partial x_i}(x,\theta) + \frac{1}{2}\sum_{i_1,i_2=1}^N \sum_{k=1}^d V_k^{i_1}(x,\theta)V_k^{i_2}(x,\theta)\frac{\partial^2 \varphi}{\partial x_{i_1}\partial x_{i_2}}(x,\theta); \\ \mathcal{L}_j\varphi(x,\theta) &= \sum_{i=1}^N V_j^i(x,\theta)\frac{\partial \varphi}{\partial x_i}(x,\theta), \quad 1 \le j \le d, \end{aligned}$$

for $(x,\theta) \in \mathbb{R}^N \times \Theta$. Application of the above differential operators is extended to vector-valued functions in the apparent way, via separate consideration of each scalar component. We denote the probability law of the process $\{X_t\}_{t\ge 0}$ under a parameter $\theta \in \Theta$ as $\mathbb{P}_\theta$, and we write $\xrightarrow{\mathbb{P}_{\theta^\dagger}}$, $\xrightarrow{\mathcal{L}_{\theta^\dagger}}$ for convergence in probability and distribution, respectively, under the true parameter $\theta^\dagger$. We write the expectation under the probability law $\mathbb{P}_\theta$ as $\mathbb{E}_\theta$ to emphasise the dependence on $\theta \in \Theta$. For $u \in \mathbb{R}^n$, $n \in \mathbb{N}$ and the multi-index $\alpha \in \{1,\dots,n\}^l$, $l \in \mathbb{N}$, we define $\partial_\alpha^u = \partial^l/\partial u_{\alpha_1}\cdots\partial u_{\alpha_l}$, i.e. an operator acting on maps $\mathbb{R}^n \to \mathbb{R}$, and then extended, by separate application on each co-ordinate, on maps $\mathbb{R}^n \to \mathbb{R}^m$, $m \in \mathbb{N}$. We denote by $C_p^\infty(\mathbb{R}^{n_1}\times\Theta;\mathbb{R}^{n_2})$, $n_1, n_2 \in \mathbb{N}$, the space of functions $f : \mathbb{R}^{n_1}\times\Theta \to \mathbb{R}^{n_2}$ such that $f(x,\theta)$ is infinitely differentiable w.r.t. $x \in \mathbb{R}^{n_1}$ for all $\theta \in \Theta$, and for any $\alpha \in \{1,\dots,n_1\}^\ell$, $\ell \ge 0$, $\partial_x^\alpha f(x,\theta)$ is of polynomial growth in $x \in \mathbb{R}^{n_1}$ uniformly in $\theta \in \Theta$. For $x, y \in \mathbb{R}^N$, we write $\langle x, y\rangle = \sum_{i=1}^N x_i y_i$ and $\|x\| = \sqrt{\langle x, x\rangle}$.

2. Hypo-elliptic SDEs

We fully specify the classes of SDEs of interest in (Hypo-I) and (Hypo-II), by providing, in each case, appropriate conditions on the collection of functionals $\{V_0, V_1, \dots, V_d\}$, motivated by modelling considerations and the existence of a Lebesgue density for the SDE transition dynamics. We illustrate later on that the imposed conditions suffice so that the locally Gaussian scheme for X_t in (Hypo-II) we put forward in this paper is non-degenerate.

2.1. Hörmander's condition

We quickly review the definition of Hörmander's condition. Consider the class of SDEs with the general form in (1). We define

$$\tilde{V}_0(x,\theta) = V_0(x,\theta) - \tfrac{1}{2}\sum_{k=1}^d \mathcal{L}_k V_k(x,\theta), \quad (x,\theta) \in \mathbb{R}^N \times \Theta.$$

From standard properties of Itô's processes, $\tilde{V}_0$ is the drift function of (1) when written as a Stratonovich-type SDE. The functionals $\{\tilde{V}_0, \dots, V_d\}$ of the Stratonovich SDE can be corresponded to differential operators, the latter applying on mappings on $\mathbb{R}^N \to \mathbb{R}^N$ and giving as outcome mappings, again, on the same spaces. In particular, we have:

$$\tilde{V}_0 \mapsto \sum_{i=1}^N \tilde{V}_0^i(x)\partial_{x_i}, \qquad V_k \mapsto \sum_{i=1}^N V_k^i(x)\partial_{x_i}, \quad 1 \le k \le d.$$

Without confusion, we use the same notation both for the SDE functionals and the corresponding differential operators. Parameter θ is removed from the expressions for simplicity. For two functionals (equivalently, differential operators) as above, $W = \sum_{i=1}^N W^i(x)\partial_{x_i}$ and $Z = \sum_{i=1}^N Z^i(x)\partial_{x_i}$, the Lie bracket is defined as

$$[W, Z] = W Z - Z W,$$

that is, for a given $x \in \mathbb{R}^N$,

$$[W, Z](x) = \sum_{i=1}^{N} \left\{ W^i(x)\partial_{x_i} Z(x) - Z^i(x)\partial_{x_i} W(x) \right\} \in \mathbb{R}^N.$$

We introduce the collections of functionals

$$\mathscr{H}_0 = \{V_1, \dots, V_d\}, \qquad \mathscr{H}_k = \Big\{ \{ [\tilde{V}_0, V], [V_r, V] \} \,:\, V \in \mathscr{H}_{k-1},\ 1 \le r \le d \Big\}, \quad k \ge 1.$$

Then, Hörmander's condition is stated as follows:

Definition. Hörmander's condition is said to hold at a point $x \in \mathbb{R}^N$ if there exists $M \ge 0$ such that

$$\mathrm{span}\Big\{ V(x) \,:\, V \in \mathscr{H}_j,\ \ 0 \le j \le M \Big\} = \mathbb{R}^N. \tag{4}$$

Hörmander's condition implies that for any $t > 0$ and an initial condition $X_0 = x \in \mathbb{R}^N$, the law of X_t is absolutely continuous w.r.t. the Lebesgue measure. Also, if the coefficients of the SDE are infinitely-times differentiable, with partial derivatives of all orders being bounded, then the Lebesgue density is smooth, see, e.g., [28,29]. When Hörmander's condition holds with $M = 0$, the condition is typically referred to as *strong* Hörmander's condition and is equivalent to the SDE being an elliptic one. It is easy to see that in our setting of a degenerate diffusion matrix, Hörmander's condition cannot hold with $M = 0$. If (4) holds with $M \ge 1$, then the term *weak* Hörmander's condition is often used, see e.g. [30]. Furthermore, if (4) holds uniformly in the initial point $x \in \mathbb{R}^N$, we have the *uniform* Hörmander's condition, which is invoked frequently in the literature, see e.g. [2,4]. Precisely, the uniform Hörmander's condition is stated as follows. There exists an integer $M \ge 0$ such that

$$\inf_{x \in \mathbb{R}^N} \ \inf_{\substack{\xi \in \mathbb{R}^N \text{ s.t.} \\ \|\xi\|=1}} \sum_{j=0}^{M} \sum_{V \in \mathscr{H}_j} \Big\langle V(x),\ \xi \Big\rangle^2 > 0. \tag{5}$$

To establish a statistical theory for models (Hypo-I) and (Hypo-II), we work below with assumptions associated with the uniform Hörmander's condition.

2.2. *Diffusion classes* (Hypo-I) *and* (Hypo-II)

We now set up separate conditions for the SDEs in classes (Hypo-I) and (Hypo-II). These will make use of the drift function, $\tilde{V}_0 = \tilde{V}_0(x, \theta)$, of the Stratonovich version of the SDEs. For $1 \le j \le k \le N$, we define the projection operator $\mathrm{proj}_{j,k} : \mathbb{R}^N \to \mathbb{R}^{k-j+1}$ as

$$x = \big[x_1, \dots, x_N\big]^\top \mapsto \mathrm{proj}_{j,k}(x) = \big[x_j, \dots, x_k\big]^\top.$$

We also introduce:

$$\widetilde{\mathscr{H}_0} = \mathscr{H}_0, \qquad \widetilde{\mathscr{H}_1} = \Big\{ [\tilde{V}_0, V] \,:\, V \in \widetilde{\mathscr{H}_0} \Big\} \qquad \widetilde{\mathscr{H}_2} = \Big\{ \big[\tilde{V}_0, [\tilde{V}_0, V]\big] \,:\, V \in \widetilde{\mathscr{H}_0} \Big\}.$$

For (Hypo-I) and (Hypo-II), we assign the following conditions to fully specify the structure of the corresponding degenerate system of SDEs.

Condition (H) (*Classes of SDEs*)**.**

I. For class (Hypo-I)*, it holds that:*

$$\begin{gathered}
\inf_{(x,\sigma) \in \mathbb{R}^N \times \Theta_\sigma} \ \inf_{\substack{\xi \in \mathbb{R}^{N_R} \\ \text{s.t. } \|\xi\|=1}} \sum_{V \in \widetilde{\mathscr{H}_0}} \Big\langle \mathrm{proj}_{N_S+1,N}\big\{V(x,\theta)\big\},\ \xi \Big\rangle^2 = \inf_{(x,\sigma) \in \mathbb{R}^N \times \Theta_\sigma} \ \inf_{\substack{\xi \in \mathbb{R}^{N_R} \\ \text{s.t. } \|\xi\|=1}} \sum_{k=1}^{d} \Big\langle V_{R,k}(x,\sigma),\ \xi \Big\rangle^2 > 0; \\
\inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} \ \inf_{\substack{\xi \in \mathbb{R}^{N} \\ \text{s.t. } \|\xi\|=1}} \sum_{j=0,1} \sum_{V \in \widetilde{\mathscr{H}_j}} \Big\langle V(x,\theta),\ \xi \Big\rangle^2 > 0.
\end{gathered} \tag{6}$$

II. In the case of class (Hypo-II)*, it holds that:*

$$\begin{gathered}
\inf_{(x,\sigma) \in \mathbb{R}^N \times \Theta_\sigma} \ \inf_{\substack{\xi \in \mathbb{R}^{N_R} \\ \text{s.t. } \|\xi\|=1}} \sum_{V \in \widetilde{\mathscr{H}_0}} \Big\langle \mathrm{proj}_{N_S+1,N}\big\{V(x,\theta)\big\},\ \xi \Big\rangle^2 = \inf_{(x,\sigma) \in \mathbb{R}^N \times \Theta_\sigma} \ \inf_{\substack{\xi \in \mathbb{R}^{N_R} \\ \text{s.t. } \|\xi\|=1}} \sum_{k=1}^{d} \Big\langle V_{R,k}(x,\sigma),\ \xi \Big\rangle^2 > 0; \\
\inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} \ \inf_{\substack{\xi \in \mathbb{R}^{N_{S_2}+N_R} \\ \text{s.t. } \|\xi\|=1}} \sum_{j=0,1} \sum_{V \in \widetilde{\mathscr{H}_j}} \Big\langle \mathrm{proj}_{N_{S_1}+1,N}\big\{V(x,\theta)\big\},\ \xi \Big\rangle^2 > 0; \\
\inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} \ \inf_{\substack{\xi \in \mathbb{R}^{N} \\ \text{s.t. } \|\xi\|=1}} \sum_{j=0,1,2} \sum_{V \in \widetilde{\mathscr{H}_j}} \Big\langle V(x,\theta),\ \xi \Big\rangle^2 > 0.
\end{gathered}$$

Note that the uniform Hörmander's condition holds for (Hypo-I) and (Hypo-II) under (H)-I and (H)-II (for $M = 1$ and $M = 2$) respectively.

Remark 1. Conditions (H)-I and (H)-II separate classes (Hypo-I) and (Hypo-II). In particular, the top equation in both (H)-I, (H)-II implies that the diffusion matrix of the rough component is of full rank, thus $X_{R,t}$ acquires the roughness of an elliptic SDE. The second equation in (H)-I, (H)-II ensures that all coordinates of component $X_{S,t}$ and $X_{S_2,t}$, respectively, possess the same smoothness as integrals of elliptic SDEs, that is, contain a Gaussian noise with the size of $\mathcal{O}(\Delta_n^{3/2})$. Finally, the third equation in (H)-II implies that component $X_{S_1,t}$ has the smoothness of second integrals of elliptic SDEs, i.e., contain a Gaussian noise of size $\mathcal{O}(\Delta_n^{5/2})$. Note, e.g., that (6) will not hold for (Hypo-II), since for the highly degenerate case, due to the drift function of the upper-most component $X_{S_1,t}$ not involving $X_{R,t}$, we have

$$\mathrm{proj}_{1,N_{S_1}}[\tilde{V}_0, V_k] = \mathbf{0}_{N_{S_1}}, \quad 1 \le k \le d.$$

To check this latter equation, notice that for both (Hypo-I) and (Hypo-II) we have: (i) V_0 and $\tilde{V}_0$ coincide on the smooth coordinates; (ii) $\tilde{V}_0 V_k$ is zero on the smooth coordinates. Then, for (Hypo-II) we additionally have that $V_k \tilde{V}_0$ is zero on the upper-most N_{S_1} coordinates due to the particular choice of $V_0 = V_0(x_{S_1}, x_{S_2})$.

Remark 2. We introduced Condition (H)-I & II upon functionals $\{V_0, V_1, \dots, V_d\}$, so that the two classes of SDEs, (Hypo-I) and (Hypo-II), possess sufficient structure to allow for their intended use for the modelling objectives in mind. It turns out that the exact same Condition (H)-I & II play a key role so that the locally Gaussian approximation for $X_{t_{i+1}} | X_{t_i} = x$ written down later in Section 3 are well-defined with a positive definite covariance matrix for all $(x, \theta) \in \mathbb{R}^N \times \Theta$.

2.2.1. An example for (H)-I

We provide an example for (H)-I via the following bivariate underdamped Langevin equation:

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= \big(-U'(q_t) - \gamma p_t\big)dt + \sigma dB_{1,t}, \end{aligned} \tag{7}$$

where $U : \mathbb{R} \to \mathbb{R}$, $\gamma > 0$ and $\sigma > 0$. We write $x = (q, p) \in \mathbb{R}^2$ and $\theta = (\gamma, \sigma) \in \Theta$, where Θ is an appropriate compact space. We have:

$$\widetilde{V}_0 = p\partial_q + (-U'(q) - \gamma p)\partial_p, \qquad V_1 = \sigma\partial_p, \qquad [\widetilde{V}_0, V_1] = -\sigma\partial_q + \gamma\sigma\partial_p,$$

and then

$$V_1(x,\sigma) = \begin{bmatrix} 0 \\ V_{R,1}(x,\sigma) \end{bmatrix} = \begin{bmatrix} 0 \\ \sigma \end{bmatrix}, \qquad [\widetilde{V}_0, V_1](x,\theta) = \begin{bmatrix} -\sigma \\ \gamma\sigma \end{bmatrix}.$$

Thus, SDE (7) satisfies Condition (H)-I and lies within the framework of class (Hypo-I).

2.2.2. An example for (H) -II

We provide an example for (H)-II via the following three-dimensional hypo-elliptic diffusion motivated from model class (QGLE-II):

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= \big(-U'(q_t) + \lambda s_t\big)dt; \\ ds_t &= \big(-\lambda p_t - \alpha s_t\big)dt + \sigma dB_{1,t}, \end{aligned} \tag{8}$$

where $U : \mathbb{R} \to \mathbb{R}$, $\alpha > 0$, $\sigma > 0$ and $\lambda \in \mathbb{R} \setminus \{0\}$. Notice that the drift function of the positional component q_t is independent of the rough component s_t. In this case, for $x = (p, q, s) \in \mathbb{R}^3$, $\theta = (\lambda, \alpha, \sigma) \in \Theta$ with some appropriate compact space Θ, we have:

$$\begin{aligned} &\tilde{V}_0 = p\,\partial_q + \big(-U'(q) + \lambda s\big)\partial_p + \big(-\lambda p - \alpha s\big)\partial_s, \qquad V_1 = \sigma\partial_s, \\ &[\tilde{V}_0, V_1] = -\lambda\sigma\partial_p + \alpha\sigma\partial_s, \qquad \big[\tilde{V}_0, [\tilde{V}_0, V_1]\big] = \lambda\sigma\partial_q - \lambda\alpha\sigma\partial_p + \sigma(-\lambda^2 + \alpha^2)\partial_s. \end{aligned}$$

We obtain:

$$V_1(x,\sigma) = \begin{bmatrix} 0 \\ 0 \\ V_R(x,\sigma) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \sigma \end{bmatrix}, \qquad [\widetilde{V}_0, V_1](x,\theta) = \begin{bmatrix} 0 \\ -\lambda\sigma \\ \alpha\sigma \end{bmatrix}, \qquad \big[\widetilde{V}_0, [\widetilde{V}_0, V_1]\big](x,\theta) = \begin{bmatrix} \lambda\sigma \\ -\lambda\alpha\sigma \\ \sigma(-\lambda^2 + \alpha^2) \end{bmatrix}.$$

Thus, SDE (8) satisfies Condition (H)-II and lies within the framework of class (Hypo-II).

3. Time-discretisation of hypo-elliptic SDEs

We discuss time-discretisation schemes for hypo-elliptic diffusions within classes (Hypo-I) and (Hypo-II), with the focus being on the performance of the schemes for the purposes of parametric inference. We set up the context by reviewing schemes proposed in literature for class (Hypo-I). We then propose a new scheme for the highly degenerate diffusion class (Hypo-II) that will later be proven to possess desirable statistical properties. Hereafter, to distinguish among the two classes of SDEs, we use the notation $X_t^{(\mathrm{I})}$ and $X_t^{(\mathrm{II})}$ for processes in (Hypo-I) and (Hypo-II), respectively.

3.1. Time-discretisation of (Hypo-I) – Brief review

We review relevant schemes for the hypo-elliptic class (Hypo-I) used in the literature. First, the classical *Euler–Maruyama* scheme is defined as follows, for $0 \le i \le n$,

$$X^{\mathrm{EM,(I)}}_{S,i+1} = X^{\mathrm{EM,(I)}}_{S,i} + V_{S,0}\big(X^{\mathrm{EM,(I)}}_{i}, \beta_S\big)\Delta_n;$$
$$X^{\mathrm{EM,(I)}}_{R,i+1} = X^{\mathrm{EM,(I)}}_{R,i} + V_{R,0}\big(X^{\mathrm{EM,(I)}}_{i}, \beta_R\big)\Delta_n + \sum_{j=1}^{d} V_{R,j}\big(X^{\mathrm{EM,(I)}}_{i}, \sigma\big) \times \big(B_{j,i+1} - B_{j,i}\big),$$

with subscript i in $X^{\mathrm{EM,(I)}}_{i}$ and $B_{j,i}$ indicating the time instance $t_i = i\Delta_n$. The approximation of the smooth component does not involve noise, thus the Euler–Maruyama scheme is degenerate. [32] studied some example bivariate SDEs with drift function $V_{S,0}(x, \beta_S) = x_R$, and showed that use of the Euler–Maruyama scheme in the high-frequency partial observation regime, where only the smooth component $X_{S,t}$ is observed, induces bias in parameter estimates. Note that in this setting the unobserved component $X_{R,t}$ is estimated via a finite-difference approach, i.e.,

$$X^{(\mathrm{I})}_{R,i} \approx \hat{X}^{(\mathrm{I})}_{R,i} = \frac{X^{(\mathrm{I})}_{S,i+1} - X^{(\mathrm{I})}_{S,i}}{\Delta_n},$$

and such an imputation is a main cause for the presence of bias in the estimation of parameter σ. To sidestep the above issue, [32] proposed the following conditionally Gaussian scheme, for $0 \le i \le n$:

$$\begin{aligned}
\widetilde{X}^{(\mathrm{I})}_{S,i+1} &= \widetilde{X}^{(\mathrm{I})}_{S,i} + V_{S,0}\big(\widetilde{X}^{(\mathrm{I})}_{i}, \beta_S\big)\Delta_n + \sum_{j=1}^{d} \mathcal{L}_j V_{S,0}(\widetilde{X}^{(\mathrm{I})}_{i}, \theta) \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} dB_{j,v}du; \\
\widetilde{X}^{(\mathrm{I})}_{R,i+1} &= \widetilde{X}^{(\mathrm{I})}_{R,i} + V_{R,0}\big(\widetilde{X}^{(\mathrm{I})}_{i}, \beta_R\big)\Delta_n + \sum_{j=1}^{d} V_{R,j}\big(\widetilde{X}^{(\mathrm{I})}_{i}, \sigma\big) \times \big(B_{j,i+1} - B_{j,i}\big).
\end{aligned} \tag{9}$$

Note that now the smooth component $\widetilde{X}^{(\mathrm{I})}_{S,i+1}$ involves Gaussian noise after application of an Itô–Taylor expansion for $V_{S,0}(X_t, \beta_S)$. Under Condition (H)-I, $(\widetilde{X}^{(\mathrm{I})}_{S,i+1}, \widetilde{X}^{(\mathrm{I})}_{R,i+1})$ is conditionally Gaussian with an invertible covariance matrix. Then, [32] utilised the well-posed likelihood of scheme (9) to estimate both the hidden paths of the rough components and the parameters via a Bayesian approach, namely Gibbs sampling. Under a high-frequency observation setting, they empirically showed that the estimate of parameter σ is asymptotically unbiased, but the estimator of the drift parameter β_R based on scheme (9) suffers from bias even in the complete observation regime.

[14] introduced the 'local Gaussian' scheme, where, for $0 \le i \le n$,

$$\begin{aligned}
\bar{X}^{(\mathrm{I})}_{S,i+1} &= \bar{X}^{(\mathrm{I})}_{S,i} + V_{S,0}\big(\bar{X}^{(\mathrm{I})}_{i}, \beta_S\big)\Delta_n + \tfrac{\Delta_n^2}{2}\mathcal{L}V_{S,0}\big(\bar{X}^{(\mathrm{I})}_{i}, \theta\big) + \sum_{j=1}^{d} \mathcal{L}_j V_{S,0}(\bar{X}^{(\mathrm{I})}_{i}, \theta) \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} dB_{j,v}du; \\
\bar{X}^{(\mathrm{I})}_{R,i+1} &= \bar{X}^{(\mathrm{I})}_{R,i} + V_{R,0}\big(\bar{X}^{(\mathrm{I})}_{i}, \beta_R\big)\Delta_n + \sum_{j=1}^{d} V_{R,j}\big(\bar{X}^{(\mathrm{I})}_{i}, \sigma\big) \times \big(B_{j,i+1} - B_{j,i}\big).
\end{aligned} \tag{LG-I}$$

Compared to (9), scheme (LG-I) includes term $\Delta_n^2(\mathcal{L}V_{S,0})(\bar{X}^{(\mathrm{I})}_{i}, \theta)/2$ in the smooth component. [14] illustrate the significance of this term for the purposes of parameter inference, by proving asymptotic consistency and normality for the contrast estimator derived from the likelihood of the discretisation scheme (LG-I), in the high-frequency, complete observation regime, namely, $n \to \infty$, $\Delta_n \to 0$, $n\Delta_n \to \infty$, under the step-size condition $\Delta_n = o(n^{-1/2})$.

Remark 3 (*[8]*)**.** Applied a strong 1.5 order scheme [22] to construct contrast estimators for class (Hypo-I) with $N_S = 1$, under strong conditions for the diffusion matrix so that the scheme becomes conditionally Gaussian. Then, they provided two separate contrast functions for estimating β_S and (β_R, σ) from the approximate Gaussian density for X_S and X_R, respectively, rather than the joint density. As noted in Remark 4.6 in [14], the separate contrast functions result in a larger asymptotic variance for the estimation for β_S compared with the single contrast estimator defined via the joint density of rough and smooth components.

3.2. Time-discretisation of (Hypo-II)

We propose a time-discretisation scheme for the second hypo-elliptic class (Hypo-II), with desirable properties for the purposes of parameter inference. The brief review of schemes for (Hypo-I) in the previous section suggests that the discretisation scheme for (Hypo-II) should satisfy the following two key criteria:

I. The scheme should be conditionally non-degenerate, i.e., the law of $X_{t_{i+1}}$ given X_{t_i} should admit a Lebesgue transition density for the full coordinates. This will allow to impute unobserved paths conditionally on observations without making use of bias-inducing finite-difference approximations.
II. The scheme should involve deterministic terms obtained from careful truncation of the stochastic Taylor expansion for the drift of the smooth component, $V_{S,0}(X_t^{(\mathrm{II})}, \beta_S)$, so that the contrast estimator corresponding to the scheme is asymptotically unbiased under the high-frequency, complete observation regime.

As for Criterion I, we will explain later in Section 4.2.1 that, indeed, use of a degenerate discretisation scheme or of finite-differences to estimate hidden components induces a bias in the estimation of parameters. Based upon the above key criteria, we propose the following discretisation scheme for (Hypo-II):

$$
\begin{aligned}
\bar{X}^{(\mathrm{II})}_{S_1,i+1} &= \mu_{S_1}\big(\Delta_n, \bar{X}^{(\mathrm{II})}_i, \theta\big) + \sum_{j=1}^{d} \mathcal{L}_j \mathcal{L} V_{S_1,0}\big(\bar{X}^{(\mathrm{II})}_i, \theta\big) \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} \int_{t_i}^{v} dB_{j,w} dv du; \\
\bar{X}^{(\mathrm{II})}_{S_2,i+1} &= \mu_{S_2}\big(\Delta_n, \bar{X}^{(\mathrm{II})}_i, \theta\big) + \sum_{j=1}^{d} \mathcal{L}_j V_{S_2,0}(\bar{X}^{(\mathrm{II})}_i, \theta) \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} dB_{j,v} du; \\
\bar{X}^{(\mathrm{II})}_{R,i+1} &= \mu_R\big(\Delta_n, \bar{X}^{(\mathrm{II})}_i, \theta\big) + \sum_{j=1}^{d} V_j\big(\bar{X}^{(\mathrm{II})}_i, \sigma\big) \times \big(B_{j,i+1} - B_{j,i}\big),
\end{aligned}
\tag{LG-II}
$$

where we have set, for $(\Delta, x, \theta) \in (0, \infty) \times \mathbb{R}^N \times \Theta$,

$$
\begin{bmatrix} \mu_{S_1}(\Delta, x, \theta) \\ \mu_{S_2}(\Delta, x, \theta) \\ \mu_R(\Delta, x, \theta) \end{bmatrix} = \begin{bmatrix} x_{S_1} + V_{S_1,0}(x_S, \beta_{S_1})\Delta + \mathcal{L} V_{S_1,0}(x, \theta)\frac{\Delta^2}{2} + \mathcal{L}^2 V_{S_1,0}(x, \theta)\frac{\Delta^3}{6} \\ x_{S_2} + V_{S_2,0}(x, \beta_{S_2})\Delta + \mathcal{L} V_{S_2,0}(x, \theta)\frac{\Delta^2}{2} \\ x_R + V_{R,0}(x, \beta_R)\Delta \end{bmatrix}.
$$

Notice that the scheme involves $3d$ Gaussian random variables:

$$
B_{j,i+1} - B_{j,i}, \quad \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} dB_{j,v} du, \quad \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} \int_{t_i}^{v} dB_{j,w} dv du, \qquad 1 \le j \le d.
$$

The latter of the above integrals appears due to the application of a third-order stochastic Taylor expansion on $V_{S_1,0}(x_S, \beta_{S_1})$, in the smoothest component $\bar{X}^{(\mathrm{II})}_{S_1,i+1}$. As we will show in Section 4, the log-likelihood based on the local Gaussian scheme (LG-II) produces a contrast estimator that is asymptotically unbiased in the high-frequency, complete observation regime. In order for the deduced contrast function to provide desirable asymptotic properties, it is required to include terms up to $\mathcal{O}(\Delta_n^3)$ in the definition of μ_{S_1}, otherwise estimation of the parameter β_R in the model (Hypo-II) can be asymptotically biased as [32] observed for some bivariate hypo-elliptic diffusions in the framework of (Hypo-I).

We denote by $\Sigma(\Delta, x, \theta)$ the covariance matrix for one-step implementation of scheme $\bar{X}^{(\mathrm{II})}$, with step-size $\Delta > 0$, current state $x \in \mathbb{R}^N$ and parameter $\theta \in \Theta$. The covariance matrix is given as:

$$
\Sigma(\Delta, x, \theta) = \begin{bmatrix} \Sigma_{S_1 S_1}(\Delta, x, \theta) & \Sigma_{S_1 S_2}(\Delta, x, \theta) & \Sigma_{S_1 R}(\Delta, x, \theta) \\ \Sigma_{S_2 S_1}(\Delta, x, \theta) & \Sigma_{S_2 S_2}(\Delta, x, \theta) & \Sigma_{S_2 R}(\Delta, x, \theta) \\ \Sigma_{R S_1}(\Delta, x, \theta) & \Sigma_{R S_2}(\Delta, x, \theta) & \Sigma_{RR}(\Delta, x, \theta) \end{bmatrix}, \tag{10}
$$

where each block matrix is specified as: for $x = (x_{S_1}, x_{S_2}, x_R) \in \mathbb{R}^N$, $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$,

$$
\begin{aligned}
&\Sigma_{S_1 S_1}(\Delta, x, \theta) \equiv \tfrac{\Delta^5}{20} a_{S_1}(x, \theta), \quad \Sigma_{S_1 S_2}(\Delta, x, \theta) \equiv \tfrac{\Delta^4}{8} \partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S, \beta_{S_1}) a_{S_2}(x, \theta); \\
&\Sigma_{S_1 R}(\Delta, x, \theta) \equiv \tfrac{\Delta^3}{6} \partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S, \beta_{S_1}) \partial_{x_R}^{\top} V_{S_2,0}(x, \beta_{S_2}) a_R(x, \sigma), \quad \Sigma_{S_2 S_1}(\Delta, x, \theta) \equiv \Sigma_{S_1 S_2}(\Delta, x, \theta)^{\top}; \\
&\Sigma_{S_2 S_2}(\Delta, x, \theta) \equiv \tfrac{\Delta^3}{3} a_{S_2}(x, \theta), \quad \Sigma_{S_2 R}(\Delta, x, \theta) \equiv \tfrac{\Delta^2}{2} \partial_{x_R}^{\top} V_{S_2,0}(x, \beta_{S_2}) a_R(x, \sigma), \quad \Sigma_{R S_1}(\Delta, x, \theta) \equiv \Sigma_{S_1 R}(\Delta, x, \theta)^{\top}; \\
&\Sigma_{R S_2}(\Delta, x, \theta) \equiv \Sigma_{S_2 R}(\Delta, x, \theta)^{\top}, \quad \Sigma_{RR}(\Delta, x, \theta) \equiv \Delta\, a_R(x, \sigma).
\end{aligned}
$$

In the above, we have set

$$
\begin{aligned}
&a_R(x, \sigma) = \sum_{k=1}^{d} V_{R,k}(x, \sigma) V_{R,k}(x, \sigma)^{\top}, \quad a_{S_2}(x, \theta) = \partial_{x_R}^{\top} V_{S_2,0}(x, \beta_{S_2})\, a_R(x, \sigma) \left(\partial_{x_R}^{\top} V_{S_2,0}(x, \beta_{S_2})\right)^{\top}; \\
&a_{S_1}(x, \theta) = \partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S, \beta_{S_1})\, a_{S_2}(x, \theta) \left(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S, \beta_{S_1})\right)^{\top}.
\end{aligned}
$$

Proposition 1. *Under Condition (H)-II, it holds that:*

$$
\inf_{(x,\sigma) \in \mathbb{R}^N \times \Theta_\sigma} |a_R(x, \sigma)| > 0, \qquad \inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} |a_{S_2}(x, \theta)| > 0, \qquad \inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} |a_{S_1}(x, \theta)| > 0. \tag{11}
$$

Then, it also holds that: for any $\Delta > 0$,

$$
\inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} |\Sigma(\Delta, x, \theta)| > 0. \tag{12}
$$

The proof is given in Appendix B. Due to Proposition 1, the covariance $\Sigma(\Delta, x, \theta)$ is invertible uniformly in x and θ for any $\Delta > 0$, thus the approximate log-likelihood based on the local Gaussian discretisation scheme (LG-II) is well-defined for the highly degenerate class (Hypo-II). We note that, in brief, result (12) follows from (11) under Condition (H)-II.

Remark 4. For elliptic diffusions, [21,36] assumed $\inf_{(x,\sigma)\in\mathbb{R}^N\times\Theta_\sigma}|a_R(x,\sigma)|>0$, and then established asymptotic properties of developed contrast estimators by introducing additional technical conditions. For the degenerate diffusion class (Hypo-II), we work with Condition (H)-II that is sufficient for (11). We make use of (11) in the proof of main results, i.e. when studying the asymptotic properties of the contrast estimator proposed later on.

4. Parameter inference for class (Hypo-II)

We explore analytically parameter inference procedures for hypo-elliptic diffusions in class (Hypo-II). We prove in Section 4.1 that a contrast estimator constructed from the conditionally Gaussian discretisation scheme (LG-II) is asymptotically unbiased under the high-frequency, complete observation regime. We illustrate the precise impact of the drift terms involved in scheme (LG-II) on the asymptotic results. In Section 4.2, we consider the partial observation regime. As observed for the case of class (Hypo-I) in the literature, we show via analytical case studies that use of finite-differences for the estimation of hidden paths leads to biased parameter estimates within (Hypo-II). Also, we explain that the local Gaussian scheme can be put into effective use within computational approaches for filtering hidden components and estimating parameters.

4.1. Complete observation regime

4.1.1. Contrast estimator

Based on the proposed scheme (LG-II) and the corresponding tractable transition density, we construct a contrast estimator for the hypo-elliptic class (Hypo-II). We write the transition density of the local Gaussian scheme (LG-II), for given $\Delta>0$, current position $x\in\mathbb{R}^N$ and parameter $\theta\in\Theta$ as:

$$y\mapsto \bar{p}_\Delta(x,y;\theta)=\frac{1}{\sqrt{(2\pi)^N\Delta^{5N_{S_1}+3N_{S_2}+N_R}|\Sigma(x,\theta)|}}\exp\Big(-\tfrac{1}{2}m(\Delta,x,y,\theta)^\top\Sigma^{-1}(x,\theta)m(\Delta,x,y,\theta)\Big),$$

where we have set, for $y=(y_{S_1},y_{S_2},y_R)\in\mathbb{R}^{N_{S_1}}\times\mathbb{R}^{N_{S_2}}\times\mathbb{R}^{N_R}$,

$$m(\Delta,x,y,\theta)=\begin{bmatrix}\frac{1}{\sqrt{\Delta^5}}\big(y_{S_1}-\mu_{S_1}(\Delta,x,\theta)\big)\\ \frac{1}{\sqrt{\Delta^3}}\big(y_{S_2}-\mu_{S_2}(\Delta,x,\theta)\big)\\ \frac{1}{\sqrt{\Delta}}\big(y_R-\mu_R(\Delta,x,\theta)\big)\end{bmatrix},\qquad \Sigma(x,\theta)=\Sigma(1,x,\theta).$$

Note that from Proposition 1, under Condition (H)-II the covariance matrix $\Sigma(x,\theta)$ is invertible for any $(x,\theta)\in\mathbb{R}^N\times\Theta$. We denote by $X_i^{(\mathrm{II})}$ the (complete) observation of a diffusion within class (Hypo-II) at time t_i, $0\le i\le n$. Then, after removing some constant terms from $-2\sum_{i=1}^n\log\bar{p}_{\Delta_n}(X_{i-1}^{(\mathrm{II})},X_i^{(\mathrm{II})};\theta)$, we define the following contrast function:

$$\ell_n(\theta)=\sum_{i=1}^n m(\Delta_n,X_{i-1}^{(\mathrm{II})},X_i^{(\mathrm{II})},\theta)^\top\,\Sigma^{-1}(X_{i-1}^{(\mathrm{II})},\theta)\,m(\Delta_n,X_{i-1}^{(\mathrm{II})},X_i^{(\mathrm{II})},\theta)+\sum_{i=1}^n\log\Big|\Sigma(X_{i-1}^{(\mathrm{II})},\theta)\Big|. \tag{13}$$

Thus, the contrast estimator for the hypo-elliptic class (Hypo-II) is defined as:

$$\hat{\theta}_n=\big(\hat{\beta}_{S_1,n},\,\hat{\beta}_{S_2,n},\,\hat{\beta}_{R,n},\,\hat{\sigma}_n\big)=\underset{\theta\in\Theta}{\mathrm{argmin}}\ \ell_n(\theta). \tag{14}$$

4.1.2. Asymptotic results

Before we state our main results, we introduce some conditions for class (Hypo-II).

- (C1) $V_j\in C_p^\infty(\mathbb{R}^N\times\Theta;\mathbb{R}^N)$, $0\le j\le d$.
- (C2) For any $1\le i\le N$, $0\le j\le d$ and $\alpha\in\{1,\dots,N\}^l$, $l\ge 0$ the following function

  $$\theta\mapsto\partial_\alpha^x V_j^i(x,\theta)$$

  is three times differentiable for all $x\in\mathbb{R}^N$. Furthermore, derivatives of the above map up to the third order are of polynomial growth in $x\in\mathbb{R}^N$ uniformly in $\theta\in\Theta$.
- (C3) The diffusion process $\{X_t\}_{t\ge0}$ defined via (Hypo-II) is ergodic under $\theta=\theta^\dagger$, with invariant distribution $\nu_{\theta^\dagger}$ on $\mathbb{R}^N$. Furthermore, all moments of $\nu_{\theta^\dagger}$ are finite.
- (C4) It holds that for all $p\ge1$, $\sup_{t>0}\mathbb{E}_{\theta^\dagger}[|X_t|^p]<\infty$.
- (C5) If it holds that

  $$V_{S,0}(x,\beta_S)=V_{S,0}(x,\beta_S^\dagger),\quad V_{R,0}(x,\beta_R)=V_{R,0}(x,\beta_R^\dagger),\quad V_R(x,\sigma)=V_R(x,\sigma^\dagger),$$

  for x in set of probability 1 under $\nu_{\theta^\dagger}$, then $\beta_S=\beta_S^\dagger$, $\beta_R=\beta_R^\dagger$, $\sigma=\sigma^\dagger$.

We write the true value of the parameter for a model in (Hypo-II) as $\theta^\dagger=(\beta_{S_1}^\dagger,\beta_{S_2}^\dagger,\beta_R^\dagger,\sigma^\dagger)\in\Theta$. The latter, $\theta^\dagger$, is assumed to lie in the interior of Θ. Recall the definition of function $V_0:\mathbb{R}^N\times\Theta\to\mathbb{R}^N$ in (3). Then, the contrast estimator defined in (14) has the following asymptotic properties in the high-frequency observation setting.

Theorem 1 (*Consistency*). *Assume that Condition (H)-II, (C1)–(C5) hold. If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then*

$$\hat{\theta}_n \xrightarrow{\mathbb{P}_{\theta^\dagger}} \theta^\dagger.$$

Theorem 2 (*Asymptotic Normality*). *Assume that Condition (H)-II, (C1)–(C5) hold. If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$ with $\Delta_n = o(n^{-1/2})$, then*

$$\begin{bmatrix} \sqrt{\frac{n}{\Delta_n^3}}\left(\hat{\beta}_{S_1,n} - \beta_{S_1}^\dagger\right) \\ \sqrt{\frac{n}{\Delta_n}}\left(\hat{\beta}_{S_2,n} - \beta_{S_2}^\dagger\right) \\ \sqrt{n\Delta_n}\left(\hat{\beta}_{R,n} - \beta_R^\dagger\right) \\ \sqrt{n}\left(\hat{\sigma}_n - \sigma^\dagger\right) \end{bmatrix} \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathcal{N}\left(\mathbf{0}_{N_\theta}, \Gamma^{-1}(\theta^\dagger)\right),$$

where the asymptotic precision matrix $\Gamma(\theta^\dagger)$ is given as:

$$\Gamma(\theta^\dagger) = \mathrm{diag}\Big(\Gamma^{\beta_{S_1}}(\theta^\dagger),\ \Gamma^{\beta_{S_2}}(\theta^\dagger),\ \Gamma^{\beta_R}(\theta^\dagger),\ \Gamma^{\sigma}(\theta^\dagger)\Big), \tag{15}$$

with the involved block matrices $\Gamma^{\beta_{S_1}}(\theta^\dagger) \in \mathbb{R}^{N_{\beta_{S_1}} \times N_{\beta_{S_1}}}$, $\Gamma^{\beta_{S_2}}(\theta^\dagger) \in \mathbb{R}^{N_{\beta_{S_2}} \times N_{\beta_{S_2}}}$, $\Gamma^{\beta_R}(\theta^\dagger) \in \mathbb{R}^{N_{\beta_R} \times N_{\beta_R}}$, $\Gamma^{\sigma}(\theta^\dagger) \in \mathbb{R}^{N_\sigma \times N_\sigma}$ specified as:

$$\begin{aligned}
\Gamma_{ij}^{\beta_{S_1}}(\theta^\dagger) &= 720 \int \partial_i^{\beta_{S_1}} V_{S_1,0}(x_S, \beta_{S_1}^\dagger)^\top a_{S_1}^{-1}(x, \theta^\dagger)\, \partial_j^{\beta_{S_1}} V_{S_1,0}(x_S, \beta_{S_1}^\dagger)\ \nu_{\theta^\dagger}(dx); \\
\Gamma_{ij}^{\beta_{S_2}}(\theta^\dagger) &= 12 \int \partial_i^{\beta_{S_2}} V_{S_2,0}(x, \beta_{S_2}^\dagger)^\top a_{S_2}^{-1}(x, \theta^\dagger)\, \partial_j^{\beta_{S_2}} V_{S_2,0}(x, \beta_{S_2}^\dagger)\ \nu_{\theta^\dagger}(dx); \\
\Gamma_{ij}^{\beta_R}(\theta^\dagger) &= \int \partial_i^{\beta_R} V_{R,0}(x, \beta_R^\dagger)^\top a_R^{-1}(x, \sigma^\dagger)\, \partial_j^{\beta_R} V_{R,0}(x, \beta_R^\dagger)\ \nu_{\theta^\dagger}(dx); \\
\Gamma_{ij}^{\sigma}(\theta^\dagger) &= \tfrac{1}{2} \int \mathrm{tr}\left(\partial_i^\sigma \Sigma(x, \theta^\dagger) \Sigma^{-1}(x, \theta^\dagger) \partial_j^\sigma \Sigma(x, \theta^\dagger) \Sigma^{-1}(x, \theta^\dagger)\right)\ \nu_{\theta^\dagger}(dx).
\end{aligned}$$

The proofs of Theorems 1 & 2 are given in Appendix C.

Remark 5. [14] proved consistency and asymptotic normality of a contrast estimator constructed via a locally Gaussian scheme for class (Hypo-I). Our proofs follow a different approach from the one in the above work. Indicatively, a condition on the step-size of $\Delta_n = o(n^{-1/2})$ is not required for our proof of consistency, whereas it is needed in [14]. We make use of preliminary convergence rates for the estimators $\hat{\beta}_{S_1,n}$, $\hat{\beta}_{S_2,n}$ and some key identities in the involved matrix calculations to control terms arising in the expansion of the logarithm of the contrast function — this method then provides the consistency for $\hat{\beta}_{R,n}$. A detailed strategy to show the consistency can be found in Appendix C.1. Also, the strategy used in our proofs can be applied to show asymptotic properties of a contrast estimator for a much wider class of highly degenerate diffusions such that Condition (H)-I or II does not hold, and higher order Lie-brackets, e.g., $[\widetilde{V}_0, [\widetilde{V}_0, [\widetilde{V}_0, V_j]]]$, are required for the uniform Hörmander's condition. For instance, let $m \geq 3$ and consider the following $(m+1)$-dimensional highly degenerate SDE:

$$\begin{aligned}
dX_{S_1,t} &= V_{S_1,0}(X_{S_2,t}, \beta_{S_1})dt; \\
dX_{S_2,t} &= V_{S_2,0}(X_{S_3,t}, \beta_{S_2})dt; \\
&\ \ \vdots \\
dX_{S_{m-1},t} &= V_{S_{m-1},0}(X_{S_m,t}, \beta_{S_{m-1}})dt; \\
dX_{S_m,t} &= V_{S_m,0}(X_{R,t}, \beta_{S_m})dt; \\
dX_{R,t} &= V_{R,0}(X_{R,t}, \beta_R)dt + V_{R,1}(X_{R,t}, \sigma)dB_{1,t},
\end{aligned}$$

where $V_{S_i,0} : \mathbb{R} \times \Theta_{\beta_{S_i}} \to \mathbb{R}$, $1 \leq i \leq m$, $V_{R,0} : \mathbb{R} \times \Theta_{\beta_R} \to \mathbb{R}$ and $V_{R,1} : \mathbb{R} \times \Theta_\sigma \to \mathbb{R}$ with some compact parameter spaces $\Theta_{\beta_{S_i}}$, Θ_{β_R} and Θ_σ. We assume a similar condition to (H) so that the uniform Hörmander's condition (5) holds with $M = m$. We provide a recipe to construct an asymptotically unbiased contrast estimator for such a general class of degenerate diffusions in Section 6. In brief, given the deduced contrast estimator, the proof of consistency will start by showing the consistency of $\hat{\beta}_{S_1,n}$ and obtaining a convergence rate. Then, utilising this result, we would next prove the consistency of $\hat{\beta}_{S_2,n}$ and get its convergence rate. We would repeat this procedure until we show the consistency of the estimators for (β_R, σ).

4.1.3. Case study – bias due to incorrect drift expansion

We have proven that the contrast estimator based on the proposed scheme (LG-II) is asymptotically unbiased under the high-frequency, complete observation regime. The inclusion of an appropriate number terms from the stochastic Taylor expansion of $V_{S_1,0}(X_{S_1,t}, X_{S_2,t}, \beta_{S_1})$ and $V_{S_2,0}(X_t, \beta_{S_2})$ in scheme (LG-II) is critical for obtaining desirable asymptotic properties. Omission of such terms will typically give rise to an asymptotic bias. In this subsection, we briefly highlight the effect of the 'drift correction' via a simple three-dimensional hypo-elliptic model from (Hypo-II). We consider the following SDE:

$$\begin{aligned}
dq_t &= p_t dt; \\
dp_t &= s_t dt; \\
ds_t &= -\beta s_t dt + \sigma dB_t,
\end{aligned} \tag{16}$$

where $\theta = (\beta, \sigma)$ is the parameter vector. We assume that all components of the system are observed, and consider the following discretisation scheme for SDE (16):

$$\bar{x}_{i+1} = \begin{bmatrix} \bar{q}_{i+1} \\ \bar{p}_{i+1} \\ \bar{s}_{i+1} \end{bmatrix} = \begin{bmatrix} \bar{q}_i + \bar{p}_i \Delta_n + \bar{s}_i \frac{\Delta_n^2}{2} \\ \bar{p}_i + \bar{s}_i \Delta_n - \beta \bar{s}_i \frac{\Delta_n^2}{2} \\ \bar{s}_i - \beta \bar{s}_i \Delta_n \end{bmatrix} + \begin{bmatrix} \sigma \times \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} \int_{t_i}^{v} dB_w dv du \\ \sigma \times \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} dB_v du \\ \sigma \times (B_{t_{i+1}} - B_{t_i}) \end{bmatrix}. \tag{17}$$

Thus, terms of size $\mathcal{O}(\Delta_n^3)$ are not included in the deterministic part of the approximation $\bar{q}_{i+1}$ of the smoothest component. Based on the conditionally Gaussian scheme (17), we define a contrast function as $(-2)\times$ (complete log-likelihood), that is, after some constants are removed:

$$\ell_n(\theta; x_{0:n}) = 6n \log \sigma + \tfrac{1}{\sigma^2} \sum_{i=1}^{n} m(\Delta_n, x_{i-1}, x_i, \theta)^\top \, \Sigma^{-1} \, m(\Delta_n, x_{i-1}, x_i, \theta),$$

where we have set $x_i = [q_i, p_i, s_i]^\top$, $0 \le i \le n$, and

$$m(\Delta_n, x_{i-1}, x_i, \theta) = \begin{bmatrix} \frac{1}{\sqrt{\Delta_n^5}} \left(q_i - q_{i-1} - p_{i-1}\Delta_n - s_{i-1} \frac{\Delta_n^2}{2} \right) \\ \frac{1}{\sqrt{\Delta_n^3}} \left(p_i - p_{i-1} - s_{i-1}\Delta_n + \beta s_{i-1} \frac{\Delta_n^2}{2} \right) \\ \frac{1}{\sqrt{\Delta_n}} \left(s_i - s_{i-1} + \beta s_{i-1} \Delta_n \right) \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \frac{1}{20} & \frac{1}{8} & \frac{1}{6} \\ \frac{1}{8} & \frac{1}{3} & \frac{1}{2} \\ \frac{1}{6} & \frac{1}{2} & 1 \end{bmatrix}.$$

Solving $\partial_\beta \ell_n(\theta; x_{0:n}) = 0$, we obtain the contrast estimator for β as $\widetilde{\beta}_n = \widetilde{g}_n / \widetilde{f}_n$, where we have defined:

$$\widetilde{f}_n = \left\{ \tfrac{1}{2} \Sigma_{32}^{-1} + \Sigma_{33}^{-1} + \tfrac{1}{4} \Sigma_{22}^{-1} + \tfrac{1}{2} \Sigma_{23}^{-1} \right\} \times \tfrac{1}{n} \sum_{i=1}^{n} s_{i-1}^2;$$

$$\widetilde{g}_n = -\frac{1}{n\sqrt{\Delta_n}} \sum_{i=1}^{n} s_{i-1} \times \left\{ \Sigma_{31}^{-1} \times \frac{q_i - q_{i-1} - p_{i-1}\Delta_n - s_{i-1}\frac{\Delta_n^2}{2}}{\sqrt{\Delta_n^5}} + \Sigma_{32}^{-1} \times \frac{p_i - p_{i-1} - s_{i-1}\Delta_n}{\sqrt{\Delta_n^3}} + \Sigma_{33}^{-1} \times \frac{s_i - s_{i-1}}{\sqrt{\Delta_n}} \right.$$
$$\left. + \tfrac{1}{2} \Sigma_{21}^{-1} \times \frac{q_i - q_{i-1} - p_{i-1}\Delta_n - s_{i-1}\frac{\Delta_n^2}{2}}{\sqrt{\Delta_n^5}} + \tfrac{1}{2} \Sigma_{22}^{-1} \times \frac{p_i - p_{i-1} - s_{i-1}\Delta_n}{\sqrt{\Delta_n^3}} + \tfrac{1}{2} \Sigma_{23}^{-1} \times \frac{s_i - s_{i-1}}{\sqrt{\Delta_n}} \right\}.$$

From the ergodicity of the process $\{s_t\}$ and Lemma 2 in the Appendix, we have that as $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$,

$$\widetilde{f}_n \xrightarrow{\mathbb{P}_{\theta^\dagger}} c_1 \times \int s^2 \, \nu_{\theta^\dagger}(ds),$$

for a non-zero constant $c_1 = \Sigma_{32}^{-1}/2 + \Sigma_{33}^{-1} + \Sigma_{22}^{-1}/4 + \Sigma_{23}^{-1}/2$, where $\nu_\theta^\dagger(ds)$ is the invariant distribution of $\{s_t\}$ under the true parameter $\theta^\dagger$. For the numerator $\widetilde{g}_n$, we apply Lemmas 2 & 3 in Appendix to obtain that

$$\widetilde{g}_n \xrightarrow{\mathbb{P}_{\theta^\dagger}} (c_1 + c_2) \times \beta^\dagger \times \int s^2 \, \nu_{\theta^\dagger}(ds),$$

for a non-zero constant $c_2 \equiv \Sigma_{31}^{-1}/6 + \Sigma_{21}^{-1}/12$. Hence, it holds that, if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$\widetilde{\beta}_n \xrightarrow{\mathbb{P}_{\theta^\dagger}} \left(1 + \tfrac{c_2}{c_1}\right) \times \beta^\dagger.$$

Thus, the drift estimation based on the discretisation scheme (17) with inappropriate drift expansion is, in general, asymptotically biased. One can check that the above bias is removed upon use of our locally Gaussian scheme (LG-II) instead of (17).

4.2. Partial observation regime

4.2.1. Case study – Bias due to finite-differences

To motivate the 'appropriateness' of the proposed locally Gaussian scheme (LG-II) in the context of parameter inference in a partial observation setting, we illustrate that naive use of finite-differences to impute hidden components (a quite common in applications) induces a bias in the estimation of the SDE parameters. To observe this, consider the model (16) again but with the drift parameter β fixed to 1:

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= s_t dt; \\ ds_t &= -s_t dt + \sigma dB_t, \end{aligned} \tag{18}$$

for $\sigma > 0$. We then apply the Euler–Maruyama scheme for the first equation of (18) and the locally Gaussian scheme (LG-I) for the remaining dynamics, i.e.,

$$\begin{bmatrix} \bar{q}_{i+1} \\ \bar{p}_{i+1} \\ \bar{s}_{i+1} \end{bmatrix} = \begin{bmatrix} \bar{q}_i + \bar{p}_i \Delta_n \\ \bar{p}_i + \bar{s}_i \Delta_n - \bar{s}_i \frac{\Delta_n^2}{2} \\ \bar{s}_i - \bar{s}_i \Delta_n \end{bmatrix} + \begin{bmatrix} 0 \\ \sigma \times \int_{t_i}^{t_{i+1}} \int_{t_i}^{u} dB_v du \\ \sigma \times \left(B_{i+1} - B_i\right) \end{bmatrix}. \tag{19}$$

Scheme (19) is degenerate since the upper-most equation does not involve noise. We now consider the estimator based on the likelihood provided by (19), given the discrete-time observations $\{q_{0:n}, s_{0:n}\}$, and with the hidden paths $p_{0:n}$ imputed via the first equation of (19) using the observations $q_{0:n}$.

Remark 6. In practice, the rough component s_t is often not observed, so one must impute the missing components $s_{0:n}$ conditionally on the observations $q_{0:n}$ by making use of the transition density (or some approximation of it) for both coordinates (p, s) of (18). One would reasonably expect that presence of bias will be typical in such a practical scenario, if it is found to be present in the simpler case when s_t is directly observed.

The complete likelihood of the discretisation scheme (19) is given as:

$$\prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{(2\pi)^2 \Delta_n^4 \sigma^4 |\Sigma|}} \exp\Big(-\frac{1}{2\sigma^2}\, m(\Delta_n, y_{i-1}, y_i)^\top\, \Sigma^{-1}\, m(\Delta_n, y_{i-1}, y_i)\Big) \times \delta\big(q_i - q_{i-1} - p_{i-1}\Delta_n\big) \right\},$$

where we have defined $y_i = [p_i, s_i]^\top$, $0 \le i \le n$, and

$$m(\Delta_n, y_{i-1}, y_i) = \begin{bmatrix} \frac{1}{\sqrt{\Delta_n^3}} \left(p_i - p_{i-1} - s_{i-1}\Delta_n + s_{i-1}\frac{\Delta_n^2}{2}\right) \\ \frac{1}{\sqrt{\Delta_n}} \left(s_i - s_{i-1} + s_{i-1}\Delta_n\right) \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{bmatrix}.$$

Integrating out $p_{0:n}$, we obtain the marginal likelihood $f_n(\sigma\,; q_{0:n+1}, s_{0:n})$ as:

$$f_n(\sigma\,; q_{0:n+1}, s_{0:n}) = \prod_{i=1}^{n} \left\{ \frac{1}{\sqrt{(2\pi)^2 \Delta_n^4 \sigma^4 |\Sigma|}} \exp\Big(-\frac{1}{2\sigma^2}\, m(\Delta_n, \hat{y}_{i-1}, \hat{y}_i)^\top\, \Sigma^{-1}\, m(\Delta_n, \hat{y}_{i-1}, \hat{y}_i)\Big) \right\},$$

where $\hat{y}_i = [\hat{p}_i, s_i]^\top$, with $\hat{p}_i = \left(q_{i+1} - q_i\right)/\Delta_n$. Then, we obtain the following contrast function for σ, after removing constant terms from $(-2) \times \log f_n(\sigma; q_{0:n+1}, s_{0:n})$:

$$\ell_n(\theta; q_{0:n+1}, s_{0:n}) = 4n \log \sigma + \frac{1}{\sigma^2} \sum_{i=1}^{n} m(\Delta_n, \hat{y}_{i-1}, \hat{y}_i)^\top\, \Sigma^{-1}\, m(\Delta_n, \hat{y}_{i-1}, \hat{y}_i). \tag{20}$$

Solving $\partial_\sigma \ell_n(\sigma; q_{0:n+1}, s_{0:n}) = 0$, we obtain the estimator $\hat{\sigma}_n$, such that:

$$(\hat{\sigma}_n)^2 = \frac{1}{2n} \sum_{i=1}^{n} m(\Delta_n, \hat{y}_{i-1}, \hat{y}_i)^\top\, \Sigma^{-1}\, m(\Delta_n, \hat{y}_{i-1}, \hat{y}_i).$$

It holds that, if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$(\hat{\sigma}_n)^2 \xrightarrow{\mathbb{P}_{\theta^\dagger}} \frac{8}{5}(\sigma^\dagger)^2. \tag{21}$$

We prove convergence (21) in Appendix E. The proof indicates that the bias arises from the higher order stochastic Taylor expansion terms of q_{i+1} which are ignored by the estimate $\hat{p}_i$. Thus, use of the finite-difference approximation for imputation of component p_t, induces an asymptotic bias at the estimation of σ.

4.2.2. Filtering and parameter inference via the proposed scheme (LG-II)

We put forward the locally Gaussian scheme (LG-II) for imputing hidden components and performing parameter inference under a partial observation regime. The scheme and its transition density on the full set of coordinates can be combined with various computational methods, e.g., Monte-Carlo Expectation Maximisation (MC-EM) and Markov Chain Monte Carlo (MCMC), similarly to earlier works [8,32] that applied some conditionally Gaussian schemes for inference of specific hypo-elliptic models within the class (Hypo-I).

We now highlight the use of a relatively straightforward Kalman filter recursion for carrying out statistical inference once the locally Gaussian scheme is adopted, for a rich sub-class of hypo-elliptic models, referred to here as *conditionally Gaussian non-linear systems*. That is, the system is originally specified as a non-linear SDE but can be treated as a linear system given components that correspond to observations. For elliptic diffusions with such a structure, continuous-time filtering and smoothing have been investigated in engineering, see e.g. Chapter 8 of [6]. Several important hypo-elliptic models used in applications fall within this sub-class, e.g., standard Langevin equations, Quasi-Markovian generalised Langevin Eq. (QGLE-I), (QGLE-II). Here, our interest lies in the sub-class derived via the general model (Hypo-II) once the constituent coefficients are specified as:

$$V_{S_1,0}(x_S, \beta_{S_1}) = C_{\beta_{S_1}} x_{S_1} + \hat{C}_{\beta_{S_1}} x_{S_2}, \quad V_{S_2,0}(x, \beta_{S_2}) = f_{S_2}(x_{S_1}, \beta_{S_1}) + C_{\beta_{S_2}} x_H;$$

$$V_{R,0}(x,\beta_R) = f_R(x_{S_1},\beta_R) + C_{\beta_R} x_H, \quad V_{R,j}(x,\sigma) = f_{R,j}(\sigma), \quad 1 \le j \le d,$$

for $x = (x_{S_1}, x_{S_2}, x_R) = (x_{S_1}, x_H) \in \mathbb{R}^{N_{S_1}} \times \mathbb{R}^{N_{S_2}+N_R}$ and $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$, where: (i) f_{S_2}, f_R, $f_{R,j}$ are vector-valued functions, allowed to be non-linear w.r.t. the state x_{S_1}; (ii) matrices

$$C_{\beta_{S_1}} \in \mathbb{R}^{N_{S_1}\times N_{S_1}}, \qquad \hat{C}_{\beta_{S_1}} \in \mathbb{R}^{N_{S_1}\times N_{S_2}}, \qquad C_{\beta_{S_2}} \in \mathbb{R}^{N_{S_2}\times (N_{S_2}+N_R)}, \qquad C_{\beta_R} \in \mathbb{R}^{N_R\times (N_{S_2}+N_R)}$$

are independent of the state x. Critically, given the observable component x_{S_1}, the drift functions are linear functions of the hidden component x_H. For the model with the above choice of coefficients, the locally Gaussian scheme (LG-II) writes as:

$$\bar{X}_{i+1} = \begin{bmatrix} \bar{X}_{S_1,i+1} \\ \bar{X}_{S_2,i+1} \\ \bar{X}_{R,i+1} \end{bmatrix} = b(\Delta_n, \bar{X}_{S_1,i}, \theta) + A(\Delta_n, \bar{X}_{S_1,i}, \theta) \begin{bmatrix} \bar{X}_{S_2,i} \\ \bar{X}_{R,i} \end{bmatrix} + w_i(\Delta_n, \theta), \tag{22}$$

for functions $b : (0,\infty)\times\mathbb{R}^{N_{S_1}}\times\Theta \to \mathbb{R}^N$, $A : (0,\infty)\times\mathbb{R}^{N_{S_1}}\times\Theta \to \mathbb{R}^{N\times(N_{S_2}+N_R)}$ and an N-dimensional Gaussian variate $w_i(\Delta_n,\theta)$. Since the right-hand side of scheme (22) is linear w.r.t. the hidden components $\bar{X}_{S_2,i}$, $\bar{X}_{R,i}$ given the observed component $\bar{X}_{S_1,i}$, one can obtain Kalman filtering and smoothing recursions, and calculate the marginal likelihood for the observations $\bar{X}_{S_1,0:n}$. We provide the closed form filtering and marginal likelihood calculations in Appendix F. We use these tools in the numerical experiments of parameter inference under the partial observation regime in Section 5 that follows.

Remark 7. [37] studied parameter inference for the QGLE of first-type in (QGLE-I), where they applied an Euler–Maruyama scheme to construct Kalman filtering and smoothing for the rough components (p_t, s_t) given the velocity p_t, with values of the latter obtained (via finite-differences) from discrete observations of the position q_t. Then, they used Kalman filtering and smoothing within an Expectation-Maximisation (EM) algorithm to estimate the parameters. However, as we have seen, such a finite-differences approach can induce bias in the estimation of the diffusion parameters.

5. Numerical studies

5.1. Linear SDE in a partial observation regime

We illustrate empirically, for an example SDE model, that parameter estimation via the proposed locally Gaussian scheme (LG-II) leads to asymptotically unbiased estimation under the partial observation regime. We also highlight the effect of the drift correction in the properties of the estimators. We again consider the model studied in Section 4.1.3, that is,

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= s_t dt; \\ ds_t &= -\beta s_t dt + \sigma dB_t, \end{aligned}$$

where $\theta = (\beta, \sigma) \in \Theta = (0,\infty)\times(0,\infty)$ is the parameter vector. In agreement with practice, we assume that only discrete observations of the smoothest component, $q_{0:n}$, are available, with an equidistant step-size Δ_n. We compute the following two estimators based on two different discretisation schemes:

$$\hat{\theta}_{n,j} = (\hat{\beta}_{n,j}, \hat{\sigma}_{n,j}) = \mathrm{argmax}_{\theta\in\Theta} \log p_j(\theta; q_{0:n}), \quad j = 1, 2,$$

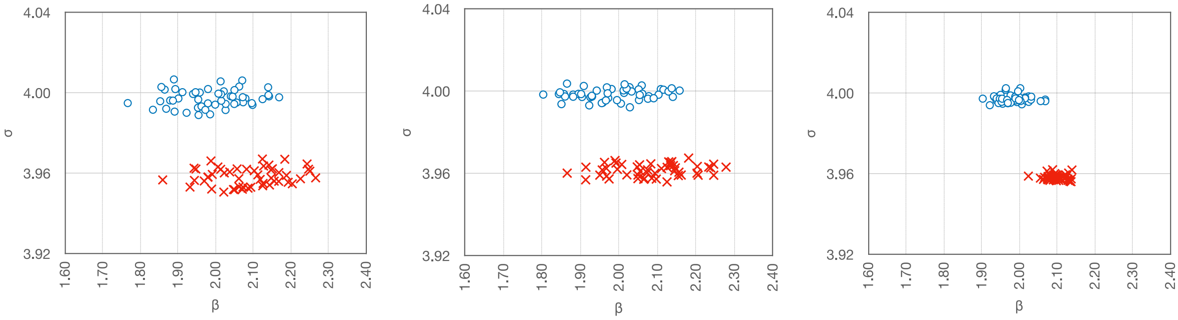
where $p_1(\theta; q_{0:n})$ is the approximate likelihood of the observations as obtained by use of Kalman filter in the setting of our locally Gaussian scheme (LG-II):

$$\begin{bmatrix} \bar{q}_{i+1} \\ \bar{p}_{i+1} \\ \bar{s}_{i+1} \end{bmatrix} = \begin{bmatrix} \bar{q}_i + \bar{p}_i\Delta_n + \bar{s}_i\frac{\Delta_n^2}{2} - \beta\bar{s}_i\frac{\Delta_n^3}{6} \\ \bar{p}_i + \bar{s}_i\Delta_n - \beta\bar{s}_i\frac{\Delta_n^2}{2} \\ \bar{s}_i - \beta\bar{s}_i\Delta_n \end{bmatrix} + \begin{bmatrix} \sigma\times\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dv du \\ \sigma\times\int_{t_i}^{t_{i+1}}\int_{t_i}^{v} dB_v du \\ \sigma\times\left(B_{i+1} - B_i\right) \end{bmatrix},$$

and $p_2(\theta; q_{0:n})$ is a different approximate likelihood obtained in the setting of the following conditionally Gaussian scheme that omits higher-order correction terms of order $\mathcal{O}(\Delta_n^2)$ and $\mathcal{O}(\Delta_n^3)$ from the stochastic Taylor expansion of the drift function of component p and q:

$$\begin{bmatrix} \widetilde{q}_{i+1} \\ \widetilde{p}_{i+1} \\ \widetilde{s}_{i+1} \end{bmatrix} = \begin{bmatrix} \widetilde{q}_i + \widetilde{p}_i\Delta_n \\ \widetilde{p}_i + \widetilde{s}_i\Delta_n \\ \widetilde{s}_i - \beta\widetilde{s}_i\Delta_n \end{bmatrix} + \begin{bmatrix} \sigma\times\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dv du \\ \sigma\times\int_{t_i}^{t_{i+1}}\int_{t_i}^{v} dB_v du \\ \sigma\times\left(B_{i+1} - B_i\right) \end{bmatrix}. \tag{23}$$

We generate 50 independent realisations of the dataset $q_{0:n}$ by sub-sampling trajectories obtained from scheme (LG-II) with a small step-size 10^{-4}. We have chosen the scheme because it is expected to have a better accuracy than other classical schemes (such as Euler–Maruyama scheme) due to the higher order stochastic Taylor expansion of drift functions. We consider the following three high-frequency scenarios for the data:



(a) Set I.   (b) Set II.   (c) Set III.

Fig. 1. Parameter estimates from 50 independent realisations of partial observations: Each blue ∘ and red × in the three figures represents one realisation of $\hat{\theta}_{n,1}$ (using (LG-II)) and $\hat{\theta}_{n,2}$ (using incorrect discretisation (23)), respectively. The true value is $\theta^{\dagger} = (\beta^{\dagger}, \sigma^{\dagger}) = (2.0, 4.0)$.

Table 3
Mean and standard deviation (in parenthesis) of $(\hat{\theta}_{n,j} - \theta^{\dagger})/\theta^{\dagger}$, $j = 1, 2$ from 50 trajectories of partial observations.

Set	Proposed scheme (LG-II)		Incorrect scheme (23)	
	$(\hat{\beta}_{n,1} - \beta^{\dagger})/\beta^{\dagger}$	$(\hat{\sigma}_{n,1} - \sigma^{\dagger})/\sigma^{\dagger}$	$(\hat{\beta}_{n,2} - \beta^{\dagger})/\beta^{\dagger}$	$(\hat{\sigma}_{n,2} - \sigma^{\dagger})/\sigma^{\dagger}$
I.	−0.0052 (0.0463)	−0.0007 (0.0011)	0.0441 (0.0477)	−0.0104 (0.0011)
II.	−0.0066 (0.0458)	−0.0003 (0.0007)	0.0438 (0.0488)	−0.0096 (0.0007)
III.	−0.0074 (0.018)	−0.0007 (0.0005)	0.0496 (0.012)	−0.0104 (0.0003)

Set I. $n = 5 \cdot 10^5$, $\Delta_n = 10^{-3}$, $T(= n\Delta_n) = 500$.
Set II. $n = 10^6$, $\Delta_n = 5 \cdot 10^{-4}$, $T = 500$.
Set III. $n = 10^7$, $\Delta_n = 10^{-3}$, $T = 10^4$.

The true parameter value is set to $\theta^{\dagger} = (\beta^{\dagger}, \sigma^{\dagger}) = (2.0, 4.0)$, and the Nelder–Mead method is applied to optimise the marginal likelihoods with the initial guess set to $(\beta_0, \sigma_0) = (3.0, 3.0)$. In Fig. 1, we plot the 50 realisations of the two different estimators. Table 3 summarises the mean and standard deviation of relative error of estimate, i.e., $(\hat{\theta}_{n,j} - \theta^{\dagger})/\theta^{\dagger}$, $j = 1, 2$, from the 50 repetitions. First, we observe that the estimates of $\hat{\theta}_{n,1}$ (using our scheme (LG-II)) are centred at the true value in all scenarios, thus in this case we have an empirical illustration of an asymptotically unbiased estimation in the partial observation setting. Secondly, it is clear from the figures and the table that the mean of estimates of $\hat{\theta}_{n,2}$ (estimator based on the conditionally Gaussian scheme without appropriate drift correction (23)) is shifted from the true value, and seems to be centred at $(\beta, \sigma) = (2.10, 3.960)$. Thus $\hat{\theta}_{n,2}$ induces an asymptotic bias in the partial observation regime, in agreement with the case study in Section 4.1.3 in the complete observation case. Notably, there is a clear separation between the two estimators of σ, and the range of realisations of $\hat{\sigma}_{n,2}$ does not cover the true value $\sigma^{\dagger} = 4.0$ for all scenarios. Comparing Set I and III in Fig. 1, we observe that the variance of drift parameter estimation decreases as T increases, and then the bias in $\hat{\beta}_{n,2}$ becomes more clear in Set III. We stress here that the bias in $\hat{\theta}_{n,2}$ is not removed with increasing n or decreasing Δ_n. Also, one can still observe the bias even if the datasets are obtained with other numerical schemes, e.g., Euler–Maruyama scheme, rather than scheme (LG-II).

5.2. Quasi-Markovian generalised Langevin equations

5.2.1. Scalar extended state

We consider the QGLE describing one-dimensional positional domain:

$$\begin{aligned} dq_t &= p_t dt; \\ dp_t &= \left(-U'(q_t) + \lambda s_t\right) dt; \\ ds_t &= (-\lambda p_t - \alpha s_t) dt + \sigma dB_t, \quad (q_0, p_0, s_0) \in \mathbb{R}^3, \end{aligned} \tag{24}$$

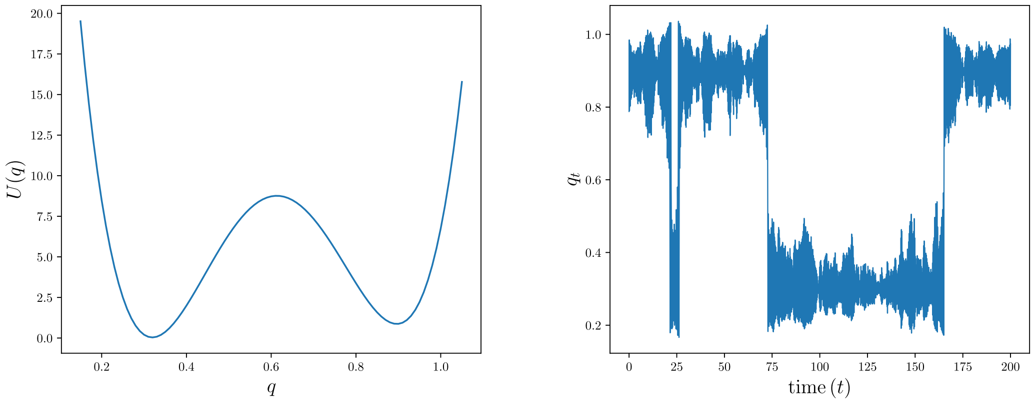
where $\alpha > 0$, $\sigma > 0$, $\lambda \in \mathbb{R} \setminus \{0\}$ and $U : \mathbb{R} \to \mathbb{R}$. In this experiment, we consider the following two choices of potential U:

$$q \mapsto U_{\mathrm{HO}}(q) = D \times \frac{q^2}{2}, \qquad q \mapsto U_{\mathrm{DW}}(q) = D \times \frac{q^2}{2} + \sin\left(\frac{1}{4} + 2q\right),$$

where $D > 0$ is a parameter. The function U_{DW} (used in experiments in the work [24]) represents an uneven double well potential, under which model (24) is non-linear. We generate 50 independent datasets by sub-sampling trajectories produced by the discretisation scheme (LG-II) with a small step-size 10^{-4} so that obtained observations correspond to $n = 2 \times 10^5$, $\Delta_n = 10^{-3}$, $T = n\Delta_n = 200$. For the complete observation regime, we compute the contrast estimator (14) for each given dataset. For experiments of partial observations, we use the trajectories of the position q_t only, extracted from the complete observations, and compute the MLE by maximising the marginal likelihood obtained from the Kalman recursion formula under the locally Gaussian scheme (LG-II), as

Table 4
Parameter estimation of the QGLE (24). Mean and standard deviation (in brackets) of $(\text{MLE} - \theta^\dagger)/\theta^\dagger$ from 50 trajectories of observations.

Potential	Parameter	Relative error of estimates	
		Complete observation	Partial observation
U_{HO}	D	−0.0000 (0.0001)	−0.0037 (0.1038)
	λ	0.0000 (0.0001)	0.0100 (0.0533)
	α	0.0057 (0.0540)	0.0091 (0.0546)
	σ	−0.0000 (0.0009)	−0.0079 (0.0524)
U_{DW}	D	0.0000 (0.0001)	−0.0027 (0.1139)
	λ	0.0000 (0.0001)	0.0201 (0.0638)
	α	0.0022 (0.0520)	0.0059 (0.0530)
	σ	0.0001 (0.0010)	−0.0166 (0.0613)



(a) Free energy $U(q)$

(b) Trajectory of q_t

Fig. 2. Left panel 2(a): The free energy used in the experiment. Right panel 2(b): A trajectory of the observable coordinate q_t from the QGLE (25).

shown in Section 4.2.2. To minimise the relevant target functions we use the adaptive moments (Adam) optimiser with the following algorithmic specifications: (step-size) = 0.2, (exponential decay rate for the first moment estimates) = 0.9, (exponential decay rate for the first moment estimates) = 0.9, (exponential decay rate for the second moment estimates) = 0.999, (additive term for numerical stability) = 10^{-8} and (number of iteration) = 2,000. The true parameters are set to $\theta^\dagger = (D^\dagger, \lambda^\dagger, \alpha^\dagger, \sigma^\dagger) = (1.0, 2.0, 4.0, 4.0)$. Also, the initial guesses for the parameter are set to $(D_0, \lambda_0, \alpha_0, \sigma_0) = (3.0, 3.0, 3.0, 3.0)$. We summarise the mean and standard deviation of $(\text{MLE} - \theta^\dagger)/\theta^\dagger$from the 50 independent trajectories in Table 4. We notice that the results for the complete observation regime are in agreement with the analytical results in Theorem 2. For instance, convergence to the true values appears to be faster for parameters (D, λ) in the smooth component p_t (recall the convergence rate $\sqrt{\Delta_n/n}$ for such parameters in the CLT of Theorem 2). Besides, under the partial observation regime, the estimates seem to be centred around the true parameter as well, with standard deviations that are larger than the ones in the case of complete observations (as expected). Thus, parameter inference carried out via the proposed locally Gaussian scheme (LG-II) appears in this case to provide unbiased estimates in the partial observation regime.

5.2.2. Multivariate extended state

We consider a QGLE with one-dimensional coordinates and multivariate extended variable, motivated from the work of Ayaz et al. [1] that studies protein-folding kinetics via a Quasi-Markovian GLE (QGLE-II) and showcases that a QGLE accurately reproduces simulations of molecular dynamics (MD) that involve memory effects in the friction. In their investigation, a one-dimensional reaction coordinate, q_t, given as the sum of the separations between native contacts, is modelled via the following QGLE:

$$\begin{aligned} dq_t &= \tfrac{1}{m} \times p_t\, dt; \\ dp_t &= -U'(q_t)dt + \sum_{l=1}^{L} s_{l,t} dt; \\ ds_{l,t} &= -\tfrac{1}{\tau_l} \times s_{l,t}\, dt - \tfrac{c_l}{\tau_l} \times p_t\, dt + \frac{\sqrt{2\beta^{-1} c_l}}{\tau_l}\, dB_{l,t}, \quad s_{l,0} \sim \mathscr{N}(0, \beta^{-1}), \quad 1 \le l \le L, \end{aligned} \tag{25}$$

where $m, \beta > 0$ denote the mass and the inverse thermal energy respectively, $\{c_l, \tau_l\}_{1 \le l \le L}$ are the unknown parameters taking positive values, for $L \ge 1$, and $U : \mathbb{R} \to \mathbb{R}$, the folding free energy landscape for proteins, is specified as $q \mapsto U(q) = -\beta^{-1} \log \nu(q)$ with $\nu(\cdot)$ being the equilibrium probability density function. QGLE (25) corresponds to the non-Markovian GLE (GLE) with the memory kernel

Table 5
Parameter estimation of the QGLE (25). Mean and standard deviation (in brackets) of $(\hat{\theta}_n - \theta^\dagger)/\theta^\dagger$ from 50 trajectories of $q_{0:n}$. The value of true parameter is $\theta^\dagger = (c_1^\dagger, \tau_1^\dagger, c_2^\dagger, \tau_2^\dagger) = (0.22, 0.007, 1.2, 4.6)$.

$(\hat{c}_{1,n} - c_1^\dagger)/c_1^\dagger$	$(\hat{\tau}_{1,n} - \tau_1^\dagger)/\tau_1^\dagger$	$(\hat{c}_{2,n} - c_2^\dagger)/c_2^\dagger$	$(\hat{\tau}_{2,n} - \tau_2^\dagger)/\tau_2^\dagger$
−0.0099 (0.0090)	0.0481 (0.0046)	−0.0364 (0.2441)	−0.0423 (0.1795)

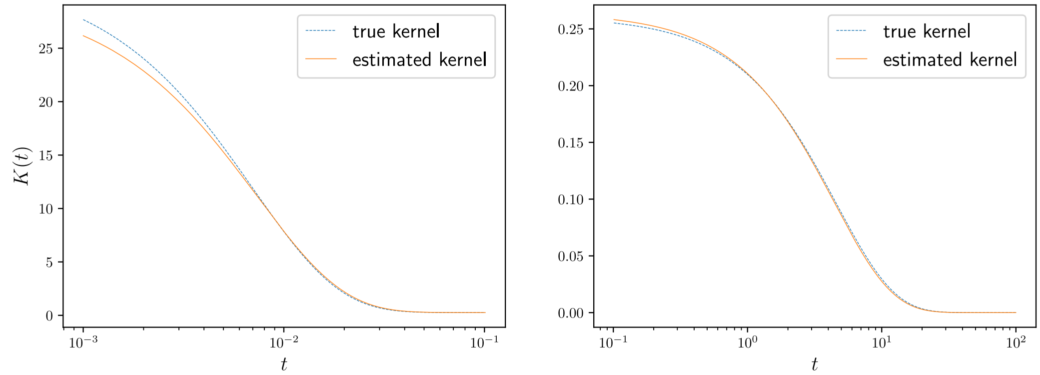


Fig. 3. Memory kernel (26) computed with the true parameter (true kernel) and with the mean value of MLEs (estimated kernel).

given as a so-called *Prony series*:

$$K(t) = \sum_{l=1}^{L} \frac{c_l}{\tau_l} \times \exp\Big(-\frac{t}{\tau_l}\Big), \quad t \geq 0. \tag{26}$$

[1] constructed QGLE (25) with $L = 5$ by determining the parameters via a least squares method so that the memory kernel (26) fits the one extracted numerically from the observed time-series of q.

In our experiment, we estimate the unknown parameters by maximising the marginal likelihood given the partial observations $q_{0:n}$. Since the mass (m) of a particle can typically be measured, we assume m to be known and set $m = 1$, in agreement with numerical experiments in the literature [8,32]. We set $L = 2$, and specify the free energy function as $q \mapsto U(q) = a(q - q_{\min})^2(q - q_{\max})^2 + bq^3$ with constants $(q_{\min}, q_{\max}, a, b) = (0.30, 0.90, 1200, 0.001)$. We set $\beta^{-1} = 2.949$ and select the true parameters as $\theta^\dagger = (c_1^\dagger, \tau_1^\dagger, c_2^\dagger, \tau_2^\dagger) = (0.22, 0.007, 1.2, 4.6)$, as such a choice closely reproduces the shape of the memory kernel estimated in [1]. We generate 50 independent trajectories of q on the time interval $[0, 1500]$ by applying scheme (LG-II) to QGLE (25) with step-size 10^{-4}. We discard the observations up to time 500 and sub-sample the datasets $q_{0:n}$ in equilibrium, so that $n = 10^6$, $\Delta_n = 10^{-3}$, $T = 1000$. In Fig. 2, we plot the shape of the free energy U and one trajectory of the component q from the QGLE (25) in the chosen setting. Notice that QGLE (25) is a conditionally Gaussian non-linear system (given the component q), thus upon adoption of the locally Gaussian discretisation (LG-II), the marginal likelihood can be calculated via the Kalman filter shown in Section 4.2.2. We use the Nelder–Mead method to optimise the marginal likelihoods with the initial value $\theta_0 = (0.1, 0.01, 1.0, 10.0)$. Table 5 summarises the mean and standard deviation of relative error of estimates $\hat{\theta}_n = (\hat{c}_{1,n}, \hat{\tau}_{1,n}, \hat{c}_{2,n}, \hat{\tau}_{2,n})$. The estimation results are overall acceptable though the estimator $\hat{\tau}_{1,n}$ is slightly biased and centred around 0.0073 — we have found empirically that such bias disappears upon choosing a smaller step-size Δ_n, e.g. $\Delta_n = 2 \times 10^{-4}$, so that $n\Delta_n^2$ becomes closer to 0. We typically observe that the standard deviation of estimates for (c_1, τ_1) is much smaller than that for (c_2, τ_2). In Fig. 3, we plot the memory kernel (26) with the parameters set equal to the mean value of the 50 MLEs to observe the level of agreement between the estimated memory kernel and the reference kernel, the latter computed with the true parameter values. The relative absolute errors between the true and estimated memory kernels are within 0.1 across periods $t \in [0.001, 10]$.

6. Conclusions and future directions

We have studied parameter inference procedures for the highly degenerate class of SDEs that includes a wide range of practical models, e.g., quasi-Markovian generalised equations (QGLEs), epidemiological models with time-varying parameters [11,34], non-linear continuous-time autoregressive models [35] and the classical Lorentz system upon consideration of noise effects [7]. We have introduced the locally Gaussian time-discretisation scheme (LG-II) and provided analytical/numerical results showcasing that parameter estimation based upon such scheme sidesteps biases that would arise under alternative schemes. The approach followed in this work for establishing our results for class (Hypo-II) are expected to also guide extensions to more general classes of degenerate diffusions, for which iterated Lie brackets of *any* order, i.e., $[\tilde{V}_0, [\tilde{V}_0, \ldots, [\tilde{V}_0, V_k]]]$, $1 \leq k \leq d$, are required for Hormander's condition to hold. Here, we draw upon the understanding obtained via the study of classes (Hypo-I) and (Hypo-II) to summarise key arguments for carrying out unbiased parameter estimation for general hypo-elliptic systems. First, in a partial observation regime, use of a degenerate discretisation (e.g. Euler–Maruyama) or equivalently of finite-differences to impute latent components will induce bias

Table 6
Size (in Δ_n) of the terms appearing in the locally Gaussian scheme (LG-II).

Component	Gaussian part	Deterministic part
$\bar{X}^{(\mathrm{II})}_{S_1,i+1}$	$\mathcal{O}(\Delta_n^{5/2})$, $\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} B_s ds du\right)$	$\mathcal{O}(\Delta_n^3)$
$\bar{X}^{(\mathrm{II})}_{S_2,i+1}$	$\mathcal{O}(\Delta_n^{3/2})$, $\left(\int_{t_i}^{t_{i+1}} B_s ds\right)$	$\mathcal{O}(\Delta_n^2)$
$\bar{X}^{(\mathrm{II})}_{R,i+1}$	$\mathcal{O}(\Delta_n^{1/2})$, $\left(B_{t_{i+1}} - B_{t_i}\right)$	$\mathcal{O}(\Delta_n)$

at estimates of diffusion coefficient parameters (recall the case study in Section 4.2.1). Also, note that finite-differences can be used only when the smooth component is determined as $dX_{S,t} = X_{R,t}dt$. Thus, a natural approach for parametric inference of degenerate diffusions is to develop a non-degenerate conditionally Gaussian scheme for the full coordinates, with the Gaussian noise obtained via high-order stochastic Taylor expansion of the drift functions. A lot of care should be given at the deterministic terms of the expansion to be included into the scheme, to avoid emergence of biases in estimates of drift parameters (recall the analytical study in Section 4.1.3 and the numerical results in Section 5.1). We summarise below the above-designated roadmap for the construction of 'correct' time-discretisation schemes for general classes of hypo-elliptic diffusions.

Step 1. For the rough component, X_R, the Euler–Maruyama scheme is applied.

Step 2. For the smooth coordinates in the model, one recursively applies stochastic Taylor expansion to drift functions so that Gaussian variates, in the form of iterated integrals involving Brownian motions, e.g. of the form $\int_{t_i}^{t_{i+1}} B_s ds$, $\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} B_s ds du$, appear in all smooth coordinates. This process is completed once the covariance-variance of the Gaussian approximation is positive definite.

Step 3. For a smooth component containing Gaussian noise of size $\mathcal{O}(\Delta_n^{(2k-1)/2})$, $k \geq 2$, the scheme should include all deterministic terms from the stochastic Taylor expansion up to size $\mathcal{O}(\Delta_n^k)$.

Indicatively, Table 6 summarises the size of deterministic and noisy parts of the locally Gaussian scheme (LG-II) for class (Hypo-II).

Our work in this paper leads to further research in several directions. In the CLT of the main analytical result for the parameter estimator (Theorem 2), the step-size Δ_n is required to satisfy $\Delta_n = o(n^{-1/2})$. An open problem for hypo-elliptic diffusions is the construction of estimators giving a CLT under a weaker condition $\Delta_n = o(n^{-1/p})$, $p \geq 3$. We expect that such a general estimator for degenerate diffusion models can be produced, with accompanying theory then following the strategy used in our proofs in this work, as we mentioned in Remark 5, that is, the proof of consistency does not require a condition of $\Delta_n = o(n^{-1/2})$. In a different direction, the effectiveness of the developed locally Gaussian scheme is yet to be studied under a low-frequency observation setting, i.e. with the step-size Δ assumed fixed and not small enough, in which case a number, say M, of inner sub-steps are introduced by the user. Under such a setting, the discretisation error of the true (intractable) density over the period of size Δ typically diminishes as M increases. In the case of elliptic diffusions, explicit rates of convergence to zero are provided in [16,19]. Finally, in this work, in the practical scenario of partial observations, we have investigated the behaviour of discretisation schemes via case studies and numerical experiments. Analytical theory would be quite instructive in this setting. Techniques used in the context of hidden Markov models (see e.g. [10]) are expected to be valuable in such a pursuit.

Funding

Yuga Iguchi acknowledges support from the Additional Funding Programme for Mathematical Sciences, delivered by EPSRC (EP/V521917/1) and the Heilbronn Institute for Mathematical Research.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Preliminaries

In Appendix A.1 we present some notation used in Appendix. In Appendix A.2 we introduce six auxiliary results needed in the proof of our main theorems (Theorems 1, 2) in Section 4.

A.1. Notation

For $0 = t_0 < \cdots < t_n$, with equi-distant step-size Δ_n, we write X_i for the observation at time t_i of the solution of the hypo-elliptic SDE (Hypo-II) under the true parameter value $\theta^\dagger$, defined upon the filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, \mathbb{P})$. We denote by $\nu_{\theta^\dagger}$ the invariant distribution of process (Hypo-II) under $\theta^\dagger$. In agreement with the structure of class (Hypo-II), we often represent $x \in \mathbb{R}^N$ and $\theta \in \Theta \subset \mathbb{R}^{N_\theta}$ as

$$x = (x_{S_1}, x_{S_2}, x_R) \in \mathbb{R}^{N_{S_1}} \times \mathbb{R}^{N_{S_2}} \times \mathbb{R}^{N_R}, \quad x_S \equiv (x_{S_1}, x_{S_2});$$

$$\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_1}} \times \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma, \quad \beta_S \equiv (\beta_{S_1}, \beta_{S_2}),$$

For $\varphi(\cdot,\theta) : \mathbb{R}^N \to \mathbb{R}$, $\theta \in \Theta$, bounded up to second derivatives, we define the differential operators $\mathcal{L}$ and $\mathcal{L}_j$, $1 \le j \le d$:

$$\mathcal{L}\varphi(x,\theta) = \sum_{i=1}^N V_0^i(x,\theta)\frac{\partial \varphi}{\partial x_i}(x,\theta) + \frac{1}{2}\sum_{i_1,i_2=1}^N \sum_{k=1}^d V_k^{i_1}(x,\theta)V_k^{i_2}(x,\theta)\frac{\partial^2 \varphi}{\partial x_{i_1}\partial x_{i_2}}(x,\theta);$$

$$\mathcal{L}_j\varphi(x,\theta) = \sum_{i=1}^N V_j^i(x,\theta)\frac{\partial \varphi}{\partial x_i}(x,\theta), \quad 1 \le j \le d.$$

Application of the above differential operators is extended to vector-valued functions in the apparent way, via separate consideration of each scalar component. We recall some notation used in the definition of the contrast function $\ell_n(\theta)$ in (13). We have that

$$\mu(\Delta,x,\theta) = \left[\mu_{S_1}(\Delta,x,\theta)^\top,\ \mu_{S_2}(\Delta,x,\theta)^\top,\ \mu_R(\Delta,x,\theta)^\top\right]^\top,$$

where

$$\begin{bmatrix} \mu_{S_1}(\Delta,x,\theta) \\ \mu_{S_2}(\Delta,x,\theta) \\ \mu_R(\Delta,x,\theta) \end{bmatrix} = \begin{bmatrix} x_{S_1} + V_{S_1,0}(x_S,\beta_{S_1})\Delta + \mathcal{L}V_{S_1,0}(x,\theta)\frac{\Delta^2}{2} + \mathcal{L}^2 V_{S_1,0}(x,\theta)\frac{\Delta^3}{6} \\ x_{S_2} + V_{S_2,0}(x,\beta_{S_2})\Delta + \mathcal{L}V_{S_2,0}(x,\theta)\frac{\Delta^2}{2} \\ x_R + V_{R,0}(x,\beta_R)\Delta \end{bmatrix}.$$

When $\Delta = 1$, we simply write

$$\mu(x,\theta) \equiv \mu(1,x,\theta).$$

For $x = (x_{S_1}, x_{S_2}, x_R) \in \mathbb{R}^N \equiv \mathbb{R}^{N_{S_1}} \times \mathbb{R}^{N_{S_2}} \times \mathbb{R}^{N_R}$, $y = (y_{S_1}, y_{S_2}, y_R) \in \mathbb{R}^N$, $\Delta > 0$ and $\theta \in \Theta$, we define

$$m(\Delta,x,y,\theta) = \left[\frac{y_{S_1}^\top - \mu_{S_1}(\Delta,x;\theta)^\top}{\sqrt{\Delta^5}},\ \frac{y_{S_2}^\top - \mu_{S_2}(\Delta,x;\theta)^\top}{\sqrt{\Delta^3}},\ \frac{y_R^\top - \mu_R(\Delta,x;\theta)^\top}{\sqrt{\Delta}}\right]^\top. \tag{27}$$

We write, for $1 \le i \le n$,

$$m_i(\Delta,\theta) \equiv m(\Delta, X_{i-1}, X_i, \theta). \tag{28}$$

We use $\Sigma(\Delta,x,\theta)$ to represent the covariance of one step of the local Gaussian scheme (LG-II) for the hypo-elliptic SDE (Hypo-II), given step-size $\Delta > 0$, initial point $x \in \mathbb{R}^N$ and parameter θ. We often write

$$\Sigma(x,\theta) \equiv \Sigma(1,x,\theta).$$

We express the inverse of $\Sigma(x,\theta)$ as:

$$\Sigma^{-1}(x,\theta) = \Lambda(x,\theta) = \begin{bmatrix} \Lambda_{S_1S_1}(x,\theta) & \Lambda_{S_1S_2}(x,\theta) & \Lambda_{S_1R}(x,\theta) \\ \Lambda_{S_2S_1}(x,\theta) & \Lambda_{S_2S_2}(x,\theta) & \Lambda_{S_2R}(x,\theta) \\ \Lambda_{RS_1}(x,\theta) & \Lambda_{RS_2}(x,\theta) & \Lambda_{RR}(x,\theta) \end{bmatrix}, \tag{29}$$

where each block matrix is specified as

$$\Lambda_{\iota_1\iota_2}(x,\theta) \in \mathbb{R}^{N_{\iota_1}\times N_{\iota_2}}, \quad \iota_1,\iota_2 \in \{S_1,S_2,R\}.$$

We emphasise here that $\Sigma(x,\theta)$ and its inverse $\Lambda(x,\theta)$ depend on x and (β_S,σ) but not on the drift parameter β_R in the rough component, and this is critical in the proof of consistency of $\hat{\beta}_{R,n}$. Thus, we sometimes write $\Sigma(x,(\beta_S,\sigma))$ and $\Lambda(x,(\beta_S,\sigma))$ to highlight the parameter dependency. We recall the definition of the matrices $a_R(x,\sigma) \in \mathbb{R}^{N_R\times N_R}$, $a_{S_1}(x,\theta) \in \mathbb{R}^{N_{S_2}\times N_{S_2}}$, $a_{S_2}(x,\theta) \in \mathbb{R}^{N_{S_1}\times N_{S_1}}$ as:

$$a_R(x,\sigma) = \sum_{k=1}^d V_{R,k}(x,\sigma)V_{R,k}(x,\sigma)^\top, \quad a_{S_2}(x,\theta) = \sum_{k=1}^d \left(\partial_{x_R}^\top V_{S_2,0}(x,\beta_{S_2})V_{R,k}(x,\sigma)\right)\left(\partial_{x_R}^\top V_{S_2,0}(x,\beta_{S_2})V_{R,k}(x,\sigma)\right)^\top;$$

$$a_{S_1}(x,\theta) = \sum_{k=1}^d \left(\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})\partial_{x_R}^\top V_{S_2,0}(x,\beta_{S_2})V_{R,k}(x,\sigma)\right)\left(\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})\partial_{x_R}^\top V_{S_2,0}(x,\beta_{S_2})V_{R,k}(x,\sigma)\right)^\top.$$

We define the mappings

$$\eta_{S_1} : \mathbb{R}^N \times \Theta_{\beta_{S_1}} \to \mathbb{R}^{N_{S_1}}, \quad \eta_{S_2} : \mathbb{R}^N \times \Theta_{\beta_{S_2}} \to \mathbb{R}^{N_{S_2}}, \quad \eta_R : \mathbb{R}^N \times \Theta_{\beta_R} \to \mathbb{R}^{N_R}$$

as

$$\eta_{S_1}(x,\beta_{S_1}) = V_{S_1,0}(x_S,\beta_{S_1}^\dagger) - V_{S_1,0}(x_S,\beta_{S_1}), \quad \eta_{S_2}(x,\beta_{S_2}) = V_{S_2,0}(x,\beta_{S_2}^\dagger) - V_{S_2,0}(x,\beta_{S_2});$$

$$\eta_R(x,\beta_R) = V_{R,0}(x,\beta_R^\dagger) - V_{R,0}(x,\beta_R).$$

We write, with a slight abuse of notation, for $0 \le i \le n$,

$$\eta_{S_1,i}(\Delta, \beta_{S_1}) \equiv \frac{\eta_{S_1}(X_i, \beta_{S_1})}{\Delta}, \quad \eta_{S_2,i}(\Delta, \beta_{S_2}) \equiv \frac{\eta_{S_2}(X_i, \beta_{S_2})}{\Delta}.$$

We denote by $\mathcal{S}$ the space of functions $f : [0,\infty)\times\mathbb{R}^N\times\Theta \to \mathbb{R}$ so that there are constants $C, q > 0$ such that $|f(\Delta, x, \theta)| \le C\Delta(1+|x|^q)$ for any $(\Delta, x, \theta) \in [0,\infty)\times\mathbb{R}^N\times\Theta$. For an $M_1 \times M_2$ matrix A, with $M_1, M_2 \ge 1$, we write each matrix entry as $[A]_{ij}$ for $1 \le i \le M_1$, $1 \le j \le M_2$. An expectation under the probability law $\mathbb{P}_\theta$ is written as $\mathbb{E}_\theta$. We write $\partial_u = \left[\partial/\partial u_1, \dots, \partial/\partial u_n\right]^\top$, $\partial_u^2 = \partial_u \partial_u^\top \equiv \left(\partial^2/\partial u_i \partial u_j\right)_{i,j=1}^n$ for the standard differential operators acting upon maps $\mathbb{R}^n \to \mathbb{R}$, $n \ge 1$. We also write $\partial_\alpha^u = \partial^l/\partial u_{\alpha_1} \cdots \partial u_{\alpha_l}$ for a multi-index $\alpha \in \{1,\dots,n\}^l$, $l \in \mathbb{N}$. For a function $g : \mathbb{R}^n \to \mathbb{R}^m$, $n, m \in \mathbb{N}$, we write:

$$\partial_u g(u)^\top = \left[\tfrac{\partial}{\partial u_i} g^j(u)\right]_{1\le i\le n,\, 1\le j\le m}, \quad \partial_u^\top g(u) = \left(\partial_u g(u)^\top\right)^\top;$$
$$\partial_\alpha^u g(u)^\top = \left[\partial_\alpha^u g^1(u), \dots, \partial_\alpha^u g^m(u)\right], \quad \partial_\alpha^u g(u) = \left(\partial_\alpha^u g(u)^\top\right)^\top.$$

A.2. Auxiliary results

Lemma 1. *Let Y_{t_i}, U be random variables, with Y_{t_i} being $\mathcal{F}_{t_i}$-measurable. If*

$$\sum_{i=1}^n \mathbb{E}_{\theta^\dagger}[Y_{t_i} \mid \mathcal{F}_{t_{i-1}}] \xrightarrow{\mathbb{P}_{\theta^\dagger}} U, \quad \sum_{i=1}^n \mathbb{E}_{\theta^\dagger}\left[(Y_{t_i})^2 \mid \mathcal{F}_{t_{i-1}}\right] \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

then $\sum_{i=1}^n Y_{t_i} \xrightarrow{\mathbb{P}_{\theta^\dagger}} U$.

Proof. See Lemma 9 in [13]. ■

Lemma 2. *Let $f : \mathbb{R}^N \times \Theta \to \mathbb{R}$ be differentiable w.r.t. $(x,\theta) \in \mathbb{R}^N \times \Theta$ with derivatives of polynomial growth in x uniformly in θ. Under conditions (C1)–(C4), it holds that, if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then*

$$\frac{1}{n}\sum_{i=1}^n f(X_{i-1}, \theta) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int f(x,\theta)\nu_{\theta^\dagger}(dx),$$

uniformly in $\theta \in \Theta$.

Proof. This is a multivariate version of Lemma 8 in [21], so we omit the proof. ■

Lemma 3. *Let $1 \le j_1, j_2 \le N$ and assume that $f : \mathbb{R}^N \times \Theta \to \mathbb{R}$ is as in Lemma 2. Under conditions (C1)–(C4), it holds that, if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then*

$$\frac{1}{n}\sum_{i=1}^n f(X_{i-1}, \theta)\, m_i^{j_1}(\Delta_n, \theta^\dagger)\, m_i^{j_2}(\Delta_n, \theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int f(x,\theta)\left[\Sigma(x,\theta^\dagger)\right]_{j_1 j_2}\nu_{\theta^\dagger}(dx); \tag{30}$$

$$\frac{1}{n\sqrt{\Delta_n}}\sum_{i=1}^n f(X_{i-1}, \theta) m_i^{j_1}(\Delta_n, \theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \tag{31}$$

uniformly in $\theta \in \Theta$.

Proof. Notice that for any $1 \le j_1, j_2 \le N$ and $0 \le i \le n$,

$$\mathbb{E}_{\theta^\dagger}\left[m_i^{j_1}(\Delta_n, \theta^\dagger) \mid \mathcal{F}_{t_{i-1}}\right] = R_{j_1}(\sqrt{\Delta_n^3}, X_{i-1}, \theta^\dagger); \tag{32}$$

$$\mathbb{E}_{\theta^\dagger}\left[m_i^{j_1}(\Delta_n, \theta^\dagger)\, m_i^{j_2}(\Delta_n, \theta^\dagger) \mid \mathcal{F}_{t_{i-1}}\right] = \left[\Sigma(X_{i-1}, \theta^\dagger)\right]_{j_1 j_2} + R_{j_1 j_2}(\Delta_n, X_{i-1}, \theta^\dagger), \tag{33}$$

where $R_{j_1 j_2}$, $R_{j_1} \in \mathcal{S}$. Applying Lemmas 1, 2 with formulae (32) and (33), we obtain (30) and (31) in the same way as in the proof of Lemma 12 in [18]. ■

Lemma 4. *Assume that Condition (H)-II holds. We have that for any $(x,\theta) \in \mathbb{R}^N \times \Theta$:*

$$\Lambda_{S_1 S_1}(x,\theta)\, \partial_{x_{S_2}}^\top V_{S_1,0}(x_S, \beta_{S_1}) = -2\Lambda_{S_1 S_2}(x,\theta).$$

Proof. We write $\Sigma(x,\theta)$, $(x,\theta) \in \mathbb{R}^N \times \Theta$, in the form of the block matrix:

$$\Sigma(x,\theta) = \begin{bmatrix} \Sigma_{S_1 S_1}(x,\theta) & \widetilde{\Sigma}(x,\theta) \\ \widetilde{\Sigma}(x,\theta)^\top & \hat{\Sigma}(x,\theta) \end{bmatrix}, \tag{34}$$

where we have set:

$$\widetilde{\Sigma}(x,\theta) = \left[\Sigma_{S_1 S_2}(x,\theta),\ \Sigma_{S_1 R}(x,\theta)\right], \quad \hat{\Sigma}(x,\theta) = \begin{bmatrix} \Sigma_{S_2 S_2}(x,\theta) & \Sigma_{S_2 R}(x,\theta) \\ \Sigma_{R S_2}(x,\theta) & \Sigma_{RR}(x,\theta) \end{bmatrix}. \tag{35}$$

Notice that under Condition (H)–II, matrix $\hat{\Sigma}(x,\theta)$ is invertible for any $(x,\theta) \in \mathbb{R}^N \times \Theta$. We write the inverse of $\hat{\Sigma}(x,\theta)$ as:

$$\hat{\Sigma}^{-1}(x,\theta) = \hat{\Lambda}(x,\theta) = \begin{bmatrix} \hat{\Lambda}_{S_2S_2}(x,\theta) & \hat{\Lambda}_{S_2R}(x,\theta) \\ \hat{\Lambda}_{RS_2}(x,\theta) & \hat{\Lambda}_{RR}(x,\theta) \end{bmatrix}.$$

Recall the notation for the inverse of $\Sigma(x,\theta)$ in . Using the inverse formula for a block matrix, we obtain:

$$\Lambda_{S_1S_2}(x,\theta) = -\Lambda_{S_1S_1}(x,\theta)\Xi(x,\theta), \tag{36}$$

where we have set

$$\Xi(x,\theta) = \Sigma_{S_1S_2}(x,\theta)\hat{\Lambda}_{S_2S_2}(x,\theta) + \Sigma_{S_1R}(x,\theta)\hat{\Lambda}_{RS_2}(x,\theta). \tag{37}$$

From the block matrix representation of $\Sigma(\Delta,x,\theta)$ in (10), we obtain

$$\Sigma_{S_1S_2}(x,\theta) = \tfrac{3}{8}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\Sigma_{S_2S_2}(x,\theta), \quad \Sigma_{S_1R}(x,\theta) = \tfrac{1}{3}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\Sigma_{S_2R}(x,\theta).$$

We then have

$$\begin{aligned} \Xi(x,\theta) &= \tfrac{3}{8}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\Sigma_{S_2S_2}(x,\theta)\hat{\Lambda}_{S_2S_2}(x,\theta) + \tfrac{1}{3}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\Sigma_{S_2R}(x,\theta)\hat{\Lambda}_{RS_2}(x,\theta) \\ &= \tfrac{1}{24}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\Sigma_{S_2S_2}(x,\theta)\hat{\Lambda}_{S_2S_2}(x,\theta) + \tfrac{1}{3}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}) \\ &= \tfrac{1}{2}\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}). \end{aligned} \tag{38}$$

In the above calculation we have used:

$$\begin{aligned} &\Sigma_{S_2S_2}(x,\theta)\hat{\Lambda}_{S_2S_2}(x,\theta) + \Sigma_{S_2R}(x,\theta)\hat{\Lambda}_{RS_2}(x,\theta) = I_{N_{S_2}\times N_{S_2}}; \\ &\hat{\Lambda}_{S_2S_2}(x,\theta) = \left(\Sigma_{S_2S_2}(x,\theta) - \Sigma_{S_2R}(x,\theta)\Sigma_{RR}^{-1}(x,\theta)\Sigma_{RS_2}(x,\theta)\right)^{-1} = 4\Sigma_{S_2S_2}^{-1}(x,\theta), \end{aligned}$$

where matrix $\Sigma_{S_2S_2}(x,\theta)$ is invertible under Condition (H)–II. Thus, from (36) and (38), we obtain

$$\Lambda_{S_1S_2}(x,\theta) = -\tfrac{1}{2}\Lambda_{S_1S_1}(x,\theta)\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}), \tag{39}$$

and the proof is now complete. ■

Lemma 5. *Assume that Condition (H)-II holds. We have that, for any $(x,\theta) \in \mathbb{R}^N \times \Theta$:*

$$\Lambda_{S_1S_1}(x,\theta) = 720\, a_{S_1}^{-1}(x,\theta); \tag{40}$$

$$\Lambda_{S_2S_2}(x,\theta) = 12a_{S_2}^{-1}(x,\theta) - \tfrac{1}{2}\Lambda_{S_2S_1}(x,\theta)\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}). \tag{41}$$

Proof. (*Proof of* (40)). First, we note that the matrices $\Sigma_{S_1S_1}(x,\theta)$, $\Sigma_{S_2S_2}(x,\theta)$, $\Sigma_{RR}(x,\theta)$ are invertible for any $(x,\theta) \in \mathbb{R}^N \times \Theta$ under Condition (H)-II. Due to the block expression of matrix $\Sigma(x,\theta)$ in (34), we have:

$$\Lambda_{S_1S_1}(x,\theta) = \left(\Sigma_{S_1S_1}(x,\theta) - \widetilde{\Sigma}(x,\theta)\hat{\Lambda}(x,\theta)\widetilde{\Sigma}(x,\theta)^{\top}\right)^{-1},$$

where $\hat{\Lambda}(x,\theta)$, the inverse of matrix $\hat{\Sigma}(x,\theta)$ given in (35), has the following block expression:

$$\hat{\Lambda}(x,\theta) = \begin{bmatrix} \hat{\Lambda}_{S_2S_2}(x,\theta) & \hat{\Lambda}_{S_2R}(x,\theta) \\ \hat{\Lambda}_{S_2R}(x,\theta)^{\top} & \hat{\Lambda}_{RR}(x,\theta) \end{bmatrix}$$

where we have set:

$$\begin{aligned} &\hat{\Lambda}_{S_2S_2}(x,\theta) = 4\Sigma_{S_2S_2}^{-1}(x,\theta), \quad \hat{\Lambda}_{S_2R}(x,\theta) = -4\Sigma_{S_2S_2}^{-1}(x,\theta)\Sigma_{S_2R}(x,\theta)\Sigma_{RR}^{-1}(x,\sigma); \\ &\hat{\Lambda}_{RR}(x,\theta) = \Sigma_{RR}^{-1}(x,\sigma) + 4\Sigma_{RR}^{-1}(x,\sigma)\Sigma_{RS_2}(x,\theta)\Sigma_{S_2S_2}^{-1}(x,\theta)\Sigma_{S_2R}(x,\theta)\Sigma_{RR}^{-1}(x,\sigma). \end{aligned}$$

We then have:

$$\hat{\Lambda}(x,\theta)\widetilde{\Sigma}(x,\theta)^{\top} = \begin{bmatrix} \tfrac{1}{2}\left(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\right)^{\top} \\ -\tfrac{1}{12}\left(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})\right)^{\top} \end{bmatrix},$$

where we used (38) for the upper block matrix, while the lower one is obtained via:

$$\begin{aligned} &\hat{\Lambda}_{S_2R}(x,\theta)^{\top}\Sigma_{S_2S_1}(x,\theta) + \hat{\Lambda}_{RR}(x,\theta)\Sigma_{RS_1}(x,\theta) \\ &= -4\Sigma_{RR}^{-1}(x,\sigma)\Sigma_{RS_2}(x,\theta)\Sigma_{S_2S_2}^{-1}(x,\theta)\Sigma_{S_2S_1}(x,\theta) + \Sigma_{RR}^{-1}(x,\sigma)\Sigma_{RS_1}(x,\theta) \\ &\qquad + 4\Sigma_{RR}^{-1}(x,\sigma)\Sigma_{RS_2}(x,\theta)\Sigma_{S_2S_2}^{-1}(x,\theta)\Sigma_{S_2R}(x,\theta)\Sigma_{RR}^{-1}(x,\sigma)\Sigma_{RS_1}(x,\theta) \\ &= -\tfrac{3}{4}\left(\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})\right)^{\top}\left(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\right)^{\top} + \tfrac{1}{6}\left(\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})\right)^{\top}\left(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\right)^{\top} \end{aligned}$$

$$
\begin{aligned}
&\quad + \tfrac{1}{2}\big(\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})\big)^{\top} \Sigma_{S_2 S_2}^{-1}(x,\theta)\Sigma_{S_2 S_2}(x,\theta)\big(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\big)^{\top} \\
&= -\tfrac{1}{12}\big(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})\big)^{\top}.
\end{aligned}
$$

Thus, we obtain:

$$
\Big(\Sigma_{S_1 S_1}(x,\theta) - \widetilde{\Sigma}(x,\theta)\hat{\Lambda}(x,\theta)\widetilde{\Sigma}(x,\theta)^{\top}\Big)^{-1} = \big(\tfrac{1}{36}\Sigma_{S_1 S_1}(x,\theta)\big)^{-1} = 720\, a_{S_1}^{-1}(x,\theta),
$$

and now the proof of (40) is complete.

(*Proof of* (41)). From the block expression of the matrix $\Sigma(x,\theta)$ in (34), we obtain:

$$
\begin{bmatrix} \Lambda_{S_2 S_2}(x,\theta) & \Lambda_{S_2 R}(x,\theta) \\ \Lambda_{R S_2}(x,\theta) & \Lambda_{RR}(x,\theta) \end{bmatrix} = \hat{\Lambda}(x,\theta) + \hat{\Lambda}(x,\theta)\widetilde{\Sigma}(x,\theta)^{\top}\Lambda_{S_1 S_1}(x,\theta)\widetilde{\Sigma}(x,\theta)\hat{\Lambda}(x,\theta).
$$

Thus, we have:

$$
\begin{aligned}
\Lambda_{S_2 S_2}(x,\theta) &= \hat{\Lambda}_{S_2 S_2}(x,\theta) + \Xi(x,\theta)^{\top}\Lambda_{S_1 S_1}(x,\theta)\Xi(x,\theta) \\
&= 4\Sigma_{S_2 S_2}^{-1}(x,\theta) + \tfrac{1}{4}\big(\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\big)^{\top}\Lambda_{S_1 S_1}(x,\theta)\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}) \\
&= 12 a_{S_2}^{-1}(x,\theta) - \tfrac{1}{2}\Lambda_{S_2 S_1}(x,\theta)\partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}),
\end{aligned}
$$

where $\Xi(x,\theta)$ is defined as in (37), and we made use of (38), (39). The proof of (41) is now complete. ■

Lemma 6. *Assume that Condition (H)-II and (C1) hold. For any $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$ and $1 \le i \le N$, we have that*

$$
\Lambda(X_{i-1},(\beta_S,\sigma))\big(m_i(\Delta_n,\theta) - m_i(\Delta_n,(\beta_S,\beta_R^{\dagger},\sigma))\big) = \sqrt{\Delta_n}\begin{bmatrix} \mathbf{0}_{N_{S_1}} \\ \mathbf{0}_{N_{S_2}} \\ b_R \end{bmatrix} + R(\sqrt{\Delta_n^3}, X_{i-1},\theta),
$$

for $\mathbb{R}^N$-valued function R with $R^j \in \mathcal{S}$, $1 \le j \le N$, where the N_R-dimensional vector b_R is specified as:

$$
b_R \equiv a_R^{-1}(X_{i-1},\sigma)\,\eta_R(X_{i-1},\beta_R).
$$

Proof. We have

$$
\begin{aligned}
m_i(\Delta_n,\theta) - m_i(\Delta_n,(\beta_S,\beta_R^{\dagger},\sigma)) &= \begin{bmatrix} \frac{\sqrt{\Delta_n}}{6}\,\partial_{x_{S_2}}^{\top} V_{S_1,0}(X_{S,i-1},\beta_{S_1})\partial_{x_R}^{\top} V_{S_2,0}(X_{i-1},\beta_{S_2})\,\eta_R(X_{i-1},\beta_R) \\ \frac{\sqrt{\Delta_n}}{2}\,\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})\,\eta_R(X_{i-1},\beta_R) \\ \sqrt{\Delta_n}\,\eta_R(X_{i-1},\beta_R) \end{bmatrix} + R(\sqrt{\Delta_n^3}, X_{i-1},\theta) \\
&= \sqrt{\Delta_n}\begin{bmatrix} \Sigma_{S_1 R}\big(X_{i-1},\theta\big) \\ \Sigma_{S_2 R}\big(X_{i-1},\theta\big) \\ \Sigma_{RR}\big(X_{i-1},\sigma\big) \end{bmatrix} a_R^{-1}(X_{i-1},\sigma)\eta_R(X_{i-1},\beta_R) + R(\sqrt{\Delta_n^3}, X_{i-1},\theta),
\end{aligned}
$$

for $\mathbb{R}^N$-valued function R with $R^j \in \mathcal{S}$, $1 \le j \le N$, where we used for $(x,\theta) \in \mathbb{R}^N \times \Theta$,

$$
\begin{aligned}
&\mathcal{L}V_{S_2,0}(x,\theta) = \partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})V_{R,0}(x,\beta_R) + v_{S_2}(x,\beta_S,\sigma); \\
&\mathcal{L}^2 V_{S_1,0}(x,\theta) = \partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1})\partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2})V_{R,0}(x,\beta_R) + v_{S_1}(x,\beta_S,\sigma),
\end{aligned}
$$

with some functions $v_{S_2} : \mathbb{R}^N \times \Theta_{\beta_S} \times \Theta_\sigma \to \mathbb{R}^{N_{S_2}}$ and $v_{S_1} : \mathbb{R}^N \times \Theta_{\beta_S} \times \Theta_\sigma \to \mathbb{R}^{N_{S_1}}$ that are independent of $\beta_R \in \Theta_{\beta_R}$. Thus, it follows that

$$
\Lambda\big(X_{i-1},(\beta_S,\sigma)\big)\big(m_i(\Delta_n,\theta) - m_i(\Delta_n,(\beta_S,\beta_R^{\dagger},\sigma))\big) = \sqrt{\Delta_n}\begin{bmatrix} b_{S_1} \\ b_{S_2} \\ b_R \end{bmatrix} a_R^{-1}(X_{i-1},\sigma)\eta_R(X_{i-1},\beta_R) + \widetilde{R}(\sqrt{\Delta_n^3}, X_{i-1},\theta),
$$

for $\widetilde{R}^j \in \mathcal{S}$, $1 \le j \le N$, where we have set:

$$
\begin{bmatrix} b_{S_1} \\ b_{S_2} \\ b_R \end{bmatrix} = \Lambda\big(X_{i-1},\theta\big)\cdot\begin{bmatrix} \Sigma_{S_1 R}\big(X_{i-1},\theta\big) \\ \Sigma_{S_2 R}\big(X_{i-1},\theta\big) \\ \Sigma_{RR}\big(X_{i-1},\sigma\big) \end{bmatrix} = \begin{bmatrix} \mathbf{0}_{N_{S_1}\times N_R} \\ \mathbf{0}_{N_{S_2}\times N_R} \\ I_{N_R\times N_R} \end{bmatrix},
$$

since it holds $\Lambda(x,\theta)\Sigma(x,\theta) = I_{N\times N}$ for each $(x,\theta) \in \mathbb{R}^N \times \Theta$. The proof is now complete. ■

Appendix B. Proof of Proposition 1

The first equation within Condition (H)-II immediately leads to:

$$\inf_{(x,\sigma)\in\mathbb{R}^N\times\Theta} |a_R(x,\sigma)| > 0.$$

Furthermore, the first and second equations of Condition (H)-II yield:

$$\inf_{(x,\theta)\in\mathbb{R}^N\times\Theta} \inf_{\substack{\xi\in\mathbb{R}^{N_{S_2}}\\ \text{s.t. } \|\xi\|=1}} \sum_{k=1}^{d} \Big\langle \mathrm{proj}_{N_{S_1}+1, N_{S_1}+N_{S_2}} \big\{ \big[\widetilde{V}_0, V_k\big](x,\theta) \big\},\ \xi \Big\rangle^2 > 0.$$

Noticing that $\mathrm{proj}_{N_{S_1}+1, N_{S_1}+N_{S_2}} \big\{ \big[\widetilde{V}_0, V_k\big](x,\theta) \big\} = \partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2}) V_{R,k}(x,\sigma)$, we obtain:

$$\inf_{(x,\theta)\in\mathbb{R}^N\times\Theta} |a_{S_2}(x,\theta)| > 0.$$

Similarly, under Condition (H)-II,

$$\inf_{(x,\theta)\in\mathbb{R}^N\times\Theta} \inf_{\substack{\xi\in\mathbb{R}^{N_{S_1}}\\ \text{s.t. } \|\xi\|=1}} \sum_{k=1}^{d} \Big\langle \mathrm{proj}_{1, N_{S_1}} \big\{ \big[\widetilde{V}_0, [\widetilde{V}_0, V_k]\big](x,\theta) \big\},\ \xi \Big\rangle^2 > 0.$$

Since $\mathrm{proj}_{1, N_{S_1}} \big\{ \big[\widetilde{V}_0, [\widetilde{V}_0, V_k]\big](x,\theta) \big\} = \partial_{x_{S_2}}^{\top} V_{S_1,0}(x_S,\beta_{S_1}) \partial_{x_R}^{\top} V_{S_2,0}(x,\beta_{S_2}) V_{R,k}(x,\sigma)$, we have that:

$$\inf_{(x,\theta)\in\mathbb{R}^N\times\Theta_\sigma} |a_{S_1}(x,\theta)| > 0.$$

Finally, from the expression of the covariance $\Sigma(\Delta,x,\theta)$ in (10), its determinant is given as:

$$|\Sigma(\Delta,x,\theta)| = \tfrac{\Delta^9}{8640} \, |a_R(x,\sigma)| \, |a_{S_1}(x,\theta)| \, |a_{S_2}(x,\theta)|.$$

Thus, (12) holds and the proof is complete.

Appendix C. Proof of main results

In this section we prove the main results, i.e. Theorems 1, 2 in Section 4 of the main text. The proofs make use of some technical results from Appendix D.

C.1. Proof of Theorem 1 – Consistency

To show consistency, we study the limit of the contrast function $\ell_n(\theta)$, defined in (13), that involves terms such as

$$\tfrac{X_{S_1,i+1}-\mu_{S_1}(\Delta_n,X_i,\theta)}{\sqrt{\Delta_n^5}}, \quad \tfrac{X_{S_2,i+1}-\mu_{S_2}(\Delta_n,X_i,\theta)}{\sqrt{\Delta_n^3}}, \quad 1\le i\le n-1,$$

where $\{X_i\}_{i=0,\dots,n}$ are discrete-time observations under the true model (Hypo-II) with parameter $\theta^\dagger$. Then, the stochastic Taylor expansion for $X_{S,i+1}$ yields

$$\begin{aligned} \tfrac{X_{S_1,i+1}-\mu_{S_1}(\Delta_n,X_i,\theta)}{\sqrt{\Delta_n^5}} &= \tfrac{V_{S_1,0}(X_{S,i},\beta_{S_1}^\dagger)-V_{S_1,0}(X_{S,i},\beta_{S_1})}{\sqrt{\Delta_n^3}} + R_{S_1}(\Delta_n,X_i,\theta); \\ \tfrac{X_{S_2,i+1}-\mu_{S_2}(\Delta_n,X_i,\theta)}{\sqrt{\Delta_n^3}} &= \tfrac{V_{S_2,0}(X_i,\beta_{S_2}^\dagger)-V_{S_2,0}(X_i,\beta_{S_2})}{\sqrt{\Delta_n}} + R_{S_2}(\Delta_n,X_i,\theta), \end{aligned} \tag{42}$$

where $R_{S_1}^{j_1}, R_{S_2}^{j_2} \in \mathcal{S}$ for $1\le j_1\le N_{S_1}$, $1\le j_2\le N_{S_2}$. Careful steps are needed to control the first terms in the right-hand sides of (42) as $\Delta_n\to 0$, within the proof of consistency. Our proof proceeds with the following strategy which extends arguments used in [18]:

Step 1. We prove consistency, along with a convergence rate, for the estimator $\hat{\beta}_{S_1,n}$. That is, if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then $\hat{\beta}_{S_1,n} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \beta_{S_1}^\dagger$. In particular, we show the rate:

$$\frac{1}{\sqrt{\Delta_n^3}} \big(\hat{\beta}_{S_1,n} - \beta_{S_1}^\dagger\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0. \tag{43}$$

Step 2. Making use of the convergence rate in (43), we prove consistency, along with a convergence rate, for the estimator $\hat{\beta}_{S_2,n}$. That is, if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then $\hat{\beta}_{S_2,n} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \beta_{S_2}^\dagger$. In particular, we show the rate:

$$\frac{1}{\sqrt{\Delta_n}} \big(\hat{\beta}_{S_2,n} - \beta_{S_2}^\dagger\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0. \tag{44}$$

Step 3. Making use of the rates in (43) and (44), we prove consistency for the estimators $(\hat{\beta}_{R,n}, \hat{\sigma}_n)$. That is, if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then $(\hat{\beta}_{R,n}, \hat{\sigma}_n) \xrightarrow{\mathbb{P}_{\theta^\dagger}} (\beta_R^\dagger, \sigma^\dagger)$.

We emphasise that in our proof of consistency, the condition $\Delta_n = o(n^{-1/2})$ is not required while [14] assumed the condition throughout the proof of consistency in the case of the degenerate diffusion class (Hypo-I). Typically, in order to show the consistency of $\hat{\beta}_{R,n}$, [14] exploited the rates of convergence

$$\sqrt{\tfrac{n}{\Delta_n}}(\beta_{S,n} - \beta_S^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad \sqrt{n}(\sigma_n - \sigma^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0$$

that are derived under the condition $\Delta_n = o(n^{-1/2})$. In contrast, in our strategy, the rates of convergence (43) and (44) are obtained without requiring $\Delta_n = o(n^{-1/2})$, and are put into effective use to avoid explosion of terms such as

$$\frac{V_{S_1,0}(X_{S,i-1}, \beta_{S_1}^\dagger) - V_{S_1,0}(X_{S,i-1}, \hat{\beta}_{S_1,n})}{\sqrt{\Delta_n^3}}, \quad \frac{V_{S_2,0}(X_i, \beta_{S_2}^\dagger) - V_{S_2,0}(X_i, \hat{\beta}_{S_2,n})}{\sqrt{\Delta_n}}$$

as $\Delta_n \to 0$ only, with the help of some results derived from straightforward matrix calculations, i.e., Lemmas 4–6 in Appendix A.2.

Remark 8. Due to (11) in Proposition 1, we have under Condition (H)-II and (C1) that:

$$a^{-1}_{S_1, i_1 j_1} \in C_p^\infty(\mathbb{R}^N \times \Theta; \mathbb{R}), \quad a^{-1}_{S_2, i_2 j_2} \in C_p^\infty(\mathbb{R}^N \times \Theta; \mathbb{R}), \quad a^{-1}_{R, i_3 j_3} \in C_p^\infty(\mathbb{R}^N \times \Theta; \mathbb{R}),$$

for $1 \le i_1, j_1 \le N_{S_1}$, $1 \le i_2, j_2 \le N_{S_2}$, and $1 \le i_3, j_3 \le N_R$. Thus, one is able to apply Lemmas 2, 3 for the (scaled) contrast function that includes the above inverse matrices in the proofs below.

C.1.1. Step 1

Consistency of the estimator $\hat{\beta}_{S_1,n}$ is deduced from the following result.

Lemma 7. *Assume that Condition (H)-II and (C1)–(C4) hold. If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then,*

$$\frac{\Delta_n^3}{n} \ell_n(\theta) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \eta_{S_1}(x, \beta_{S_1})^\top \Lambda_{S_1 S_1}(x, \theta) \eta_{S_1}(x, \beta_{S_1}) \nu_{\theta^\dagger}(dx), \tag{45}$$

uniformly in $\theta \in \Theta$.

The proof is given in Appendix D.1. Lemma 7 implies the consistency of $\hat{\beta}_{S_1,n}$ via the following discussion. We write the right-hand side of (45) as $\mathscr{Q}(\theta)$, $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$. Under Condition (H)-II, it follows from Proposition 1 and (40) in Lemma 5 that for all non-zero $\xi \in \mathbb{R}^{N_{S_1}}$,

$$\inf_{(x,\theta) \in \mathbb{R}^N \times \Theta} \xi^\top \Lambda_{S_1 S_1}(x, \theta) \xi > 0.$$

Thus, it holds under the identifiability condition (C5) that for any $\varepsilon > 0$ there exists a constant $\delta > 0$ such that

$$\mathbb{P}_{\theta^\dagger}\left(\left| \hat{\beta}_{S_1,n} - \beta_{S_1}^\dagger \right| > \varepsilon \right) \le \mathbb{P}_{\theta^\dagger}\left(\left| \mathscr{Q}(\hat{\theta}_n) \right| > \delta \right).$$

Then, Lemma 7 together with the definition of the estimator $\hat{\theta}_n$ yields:

$$\begin{aligned} \mathbb{P}_{\theta^\dagger}\left(\left| \mathscr{Q}(\hat{\theta}_n) \right| > \delta \right) &\le \mathbb{P}_{\theta^\dagger}\left(\left| \tfrac{\Delta_n^3}{n} \ell_n(\beta_{S_1}^\dagger, \hat{\beta}_{S_2,n}, \hat{\beta}_{R,n}, \hat{\sigma}_n) - \tfrac{\Delta_n^3}{n} \ell_n(\hat{\theta}_n) + \mathscr{Q}(\hat{\theta}_n) \right| > \delta \right) \\ &\le \mathbb{P}_{\theta^\dagger}\left(\sup_{\theta \in \Theta} \left| \tfrac{\Delta_n^3}{n} \ell_n(\beta_{S_1}^\dagger, \beta_{S_2}, \beta_R, \sigma_n) - \tfrac{\Delta_n^3}{n} \ell_n(\theta) + \mathscr{Q}(\theta) \right| > \delta \right) \to 0 \end{aligned}$$

as $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, which leads to the consistency of $\hat{\beta}_{S_1,n}$.

We now prove the rate of convergence in (43). Considering the Taylor expansion of $\partial_{\beta_{S_1}} \ell_n(\hat{\theta}_n)$ around the derivative $\partial_{\beta_{S_1}} \ell_n(\beta_{S_1}^\dagger, \hat{\beta}_{S_2,n}, \hat{\beta}_{R,n}, \hat{\sigma}_n)$ with an appropriate scaling factor, we obtain

$$\mathscr{A}_{S_1,n}(\beta_{S_1}^\dagger, \hat{\beta}_{S_2,n}, \hat{\beta}_{R,n}, \hat{\sigma}_n) = \mathscr{B}_{S_1,n}(\hat{\theta}_n) \times \frac{1}{\sqrt{\Delta_n^3}}(\hat{\beta}_{S_1,n} - \beta_{S_1}^\dagger),$$

where we have set, for $\theta = (\beta_{S_1}, \theta^{-\beta_{S_1}}) \in \Theta$ with $\theta^{-\beta_{S_1}} \equiv (\beta_{S_2}, \beta_R, \sigma)$,

$$\mathscr{A}_{S_1,n}(\theta) = -\frac{\sqrt{\Delta_n^3}}{n} \partial_{\beta_{S_1}} \ell_n(\theta), \quad \mathscr{B}_{S_1,n}(\theta) = \frac{\Delta_n^3}{n} \int_0^1 \partial^2_{\beta_{S_1}} \ell_n\left(\beta_{S_1}^\dagger + \lambda(\beta_{S_1} - \beta_{S_1}^\dagger), \theta^{-\beta_{S_1}}\right) d\lambda.$$

For the matrix $\mathscr{B}_{S_1,n}(\theta)$, we have the following result:

Lemma 8. *Assume that Condition (H)-II and (C1)–(C5) hold. If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then*

$$\mathscr{B}_{S_1,n}(\hat{\beta}_{S_1,n}, \theta^{-\beta_{S_1}}) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \partial_{\beta_{S_1}} \left(V_{S_1,0}(x_S, \beta_{S_1}^\dagger) \right)^\top \Lambda_{S_1 S_1}\left(x, (\beta_{S_1}^\dagger, \theta^{-\beta_{S_1}})\right) \partial^\top_{\beta_{S_1}} V_{S_1,0}(x_S, \beta_{S_1}^\dagger) \nu_{\theta^\dagger}(dx),$$

uniformly in $\theta^{-\beta_{S_1}} \equiv (\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma$.

We give the proof in Appendix D.2. We now check that $\mathscr{A}_{S_1,n}(\beta_{S_1}^{\dagger},\theta^{-\beta_{S_1}}) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \mathbf{0}_{N_{\beta_{S_1}}}$ as $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, uniformly in $\theta^{-\beta_{S_1}}=(\beta_{S_2},\beta_R,\sigma)$. We have:

$$\mathscr{A}_{S_1,n}(\beta_{S_1}^{\dagger},\theta^{-\beta_{S_1}}) = \widetilde{\mathscr{A}}_{S_1,n}(\beta_{S_1}^{\dagger},\theta^{-\beta_{S_1}}) + \frac{1}{n}\sum_{i=1}^{n} R\big(\sqrt{\Delta_n}, X_{i-1}, (\beta_{S_1}^{\dagger},\theta^{-\beta_{S_1}})\big),$$

for $R^j\in\mathcal{S}$, $1\le j\le N_{\beta_{S_1}}$, where we have set, for $\theta=(\beta_{S_1},\beta_{S_2},\beta_R,\sigma)\in\Theta$,

$$\widetilde{\mathscr{A}}_{S_1,n}(\theta) = \frac{1}{n\sqrt{\Delta_n}}\sum_{i=1}^{n}\partial_{\beta_{S_1}}\big(V_{S_1,0}(X_{S,i-1},\beta_{S_1})\big)^{\top}\boldsymbol{\Phi}(X_{i-1},\theta)\,\eta_{S_2}(X_{i-1},\beta_{S_2})$$

with $\boldsymbol{\Phi}:\mathbb{R}^N\times\Theta\to\mathbb{R}^{N_{S_1}\times N_{S_2}}$ defined as:

$$\boldsymbol{\Phi}(x,\theta) = \Lambda_{S_1S_1}(x,\theta)\,\partial_{x_{S_2}}^{\top}V_{S_1,0}(x_S,\beta_{S_1}) + 2\Lambda_{S_1S_2}(x,\theta). \tag{46}$$

From Lemmas 2 and 3 in Appendix A, we immediately have that if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then

$$\frac{1}{n}\sum_{i=1}^{n} R\big(\sqrt{\Delta_n}, X_{i-1}, (\beta_{S_1}^{\dagger},\,\theta^{-\beta_{S_1}})\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \mathbf{0}_{N_{\beta_{S_1}}},$$

uniformly in $\theta^{-\beta_{S_1}}$. Furthermore, we have from Lemma 4 that, for any $\theta\in\Theta$, $\widetilde{\mathscr{A}}_{S_1,n}(\theta)=\mathbf{0}_{N_{\beta_{S_1}}}$ with probability 1, since $\boldsymbol{\Phi}(x,\theta)=\mathbf{0}_{N_{S_1}\times N_{S_2}}$ for any $(x,\theta)\in\mathbb{R}^N\times\Theta$. Hence, we obtain $\mathscr{A}_{S_1,n}(\beta_{S_1}^{\dagger},\theta^{-\beta_{S_1}}) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \mathbf{0}_{N_{\beta_{S_1}}}$, and now convergence (43) holds.

C.1.2. Step 2

Making use of convergence (43), we obtain the following result whose proof is postponed to Appendix D.3.

Lemma 9. *Assume that Condition (H)-II and (C1)–(C5) hold. If $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then*

$$\frac{\Delta_n}{n}\ell_n(\hat\beta_{S_1,n},\beta_{S_2},\beta_R,\sigma)\xrightarrow{\mathbb{P}_{\theta^\dagger}} 12\times\int \eta_{S_2}(x,\beta_{S_2})^{\top}a_{S_2}^{-1}\big(x,(\beta_{S_1}^{\dagger},\beta_{S_2},\sigma)\big)\,\eta_{S_2}(x,\beta_{S_2})\,\nu_{\theta^\dagger}(dx),$$

uniformly in $(\beta_{S_2},\beta_R,\sigma)\in\Theta_{\beta_{S_2}}\times\Theta_{\beta_R}\times\Theta_{\sigma}$.

This result leads to the consistency of $\hat\beta_{S_2,n}$ following an argument similar to the one used in **Step 1** to show consistency of $\hat\beta_{S_1,n}$. To prove convergence (44), we apply a Taylor expansion on the contrast function to get:

$$\mathscr{A}_{S,n}(\beta_{S_1}^{\dagger},\beta_{S_2}^{\dagger},\hat\beta_{R,n},\hat\sigma_n) = \mathscr{B}_{S,n}(\hat\theta_n)\times\begin{bmatrix}\frac{1}{\sqrt{\Delta_n^3}}(\hat\beta_{S_1,n}-\beta_{S_1}^{\dagger})\\ \frac{1}{\sqrt{\Delta_n}}(\hat\beta_{S_2,n}-\beta_{S_2}^{\dagger})\end{bmatrix},$$

where we have set, for $\theta=(\beta_S,\beta_R,\sigma)\in\Theta$ with $\beta_S\equiv(\beta_{S_1},\beta_{S_2})$,

$$\mathscr{A}_{S,n}(\theta)=\begin{bmatrix}-\frac{\sqrt{\Delta_n^3}}{n}\partial_{\beta_{S_1}}\ell_n(\theta)\\ -\frac{\sqrt{\Delta_n}}{n}\partial_{\beta_{S_2}}\ell_n(\theta)\end{bmatrix},\qquad \mathscr{B}_{S,n}(\theta)=\int_0^1 M_{\beta_S,n}\,\partial_{\beta_S}^2\ell_n\big(\beta_S^{\dagger}+\lambda(\beta_S-\beta_S^{\dagger}),\beta_R,\sigma\big)\,M_{\beta_S,n}d\lambda,$$

and $M_{\beta_S,n}\in\mathbb{R}^{N_{\beta_S}\times N_{\beta_S}}$ is defined as:

$$M_{\beta_S,n}=\mathrm{diag}\Big(\big[\underbrace{\sqrt{\tfrac{\Delta_n^3}{n}},\dots,\sqrt{\tfrac{\Delta_n^3}{n}}}_{N_{\beta_{S_1}}},\ \underbrace{\sqrt{\tfrac{\Delta_n}{n}},\dots,\sqrt{\tfrac{\Delta_n}{n}}}_{N_{\beta_{S_2}}}\big]^{\top}\Big).$$

Convergence (44) is immediately deduced from the following result.

Lemma 10. *Assume that Condition (H)-II and (C1)–(C5) hold. If $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then*

$$\mathscr{A}_{S,n}(\beta_S^{\dagger},\beta_R,\sigma)\xrightarrow{\mathbb{P}_{\theta^\dagger}}\mathbf{0}_{N_{\beta_S}}; \tag{47}$$

$$\mathscr{B}_{S,n}\big(\hat\beta_{S,n},\beta_R,\sigma\big)\xrightarrow{\mathbb{P}_{\theta^\dagger}} 2\times\mathrm{diag}\Big(\mathscr{B}_{S_1S_1}(\beta_S^{\dagger},\beta_R,\sigma),\ \mathscr{B}_{S_2S_2}(\beta_S^{\dagger},\beta_R,\sigma)\Big), \tag{48}$$

uniformly in $(\beta_R,\sigma)\in\Theta_{\beta_R}\times\Theta_\sigma$, where $\beta_S^{\dagger}\equiv(\beta_{S_1}^{\dagger},\beta_{S_2}^{\dagger})$, $\hat\beta_{S,n}\equiv(\hat\beta_{S_1,n},\hat\beta_{S_2,n})$ and we have set:

$$\mathscr{B}_{S_1S_1}(\theta)=720\int\partial_{\beta_{S_1}}\big(V_{S_1,0}(x_S,\beta_{S_1})\big)^{\top}a_{S_1}^{-1}(x,\theta)\,\partial_{\beta_{S_1}}^{\top}V_{S_1,0}(x_S,\beta_{S_1})\,\nu_{\theta^\dagger}(dx);$$

$$\mathscr{B}_{S_2S_2}(\theta)=12\int\partial_{\beta_{S_2}}\big(V_{S_2,0}(x,\beta_{S_2})\big)^{\top}a_{S_2}^{-1}(x,\theta)\partial_{\beta_{S_2}}^{\top}V_{S_2,0}(x,\beta_{S_2})\,\nu_{\theta^\dagger}(dx),$$

for $x \in \mathbb{R}^N$, $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$, *where* $\beta_S = (\beta_{S_1}, \beta_{S_2})$.

We give the proof in Appendix D.4.

C.1.3. Step 3

Finally, we prove the consistency of estimators $(\hat{\beta}_{R,n}, \hat{\sigma}_n)$. Working with the rates of convergence (43) and (44), we obtain the following result leading to the consistency of $\hat{\sigma}_n$:

Lemma 11. *Assume that Condition (H)-II and (C1)–(C5) hold. If* $n \to \infty$, $\Delta_n \to 0$ *and* $n\Delta_n \to \infty$, *then*

$$\frac{1}{n}\ell_n\big(\hat{\beta}_{S,n}, \beta_R, \sigma\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \Big\{ \mathrm{tr}\big(\Lambda(x, (\beta_S^\dagger, \sigma)) \Sigma(x, (\beta_S^\dagger, \sigma^\dagger))\big) + \log \Big|\Sigma(x, (\beta_S^\dagger, \sigma))\Big| \Big\} \nu_{\theta^\dagger}(dx),$$

uniformly in $(\beta_R, \sigma) \in \Theta_{\beta_R} \times \Theta_\sigma$.

We provide the proof in Appendix D.5. To show the consistency of $\hat{\beta}_{R,n}$, we consider, for $\beta_R \in \Theta_{\beta_R}$,

$$\mathscr{L}(\beta_R) := \frac{1}{n\Delta_n}\ell_n(\hat{\beta}_{S_1,n}, \hat{\beta}_{S_2,n}, \beta_R, \hat{\sigma}_n) - \frac{1}{n\Delta_n}\ell_n(\hat{\beta}_{S_1,n}, \hat{\beta}_{S_2,n}, \beta_R^\dagger, \hat{\sigma}_n).$$

The consistency of estimator $\hat{\beta}_{R,n}$ is obtained via the following result whose proof is given in Appendix D.6.

Lemma 12. *Assume that Condition (H)-II and (C1)–(C5) hold. If* $n \to \infty$, $\Delta_n \to 0$ *and* $n\Delta_n \to \infty$, *then*

$$\mathscr{L}(\beta_R) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \eta_R(x, \beta_R)^\top a_R^{-1}\big(x, \sigma^\dagger\big)\, \eta_R(x, \beta_R)\, \nu_{\theta^\dagger}(dx),$$

uniformly in $\beta_R \in \Theta_{\beta_R}$.

The proof of consistency for the contrast estimator $\hat{\theta}_n$ is now complete.

C.2. Proof of Theorem 2 – asymptotic normality

We consider the Taylor expansion of the contrast function $\ell_n(\theta)$:

$$\mathscr{C}_n(\theta^\dagger) = \int_0^1 \mathscr{I}_n\big(\theta^\dagger + \lambda(\hat{\theta}_n - \theta^\dagger)\big) d\lambda \times M_n(\hat{\theta}_n - \theta^\dagger)$$

where we have set, for $\theta \in \Theta$,

$$\mathscr{C}_n(\theta) = -M_n^{-1}\,\partial_\theta \ell_n(\theta), \quad \mathscr{I}_n(\theta) = M_n^{-1}\,\partial_\theta^2 \ell_n(\theta)\, M_n^{-1}, \quad M_n = \mathrm{diag}(v_n),$$

with the N_θ-dimensional vector v_n defined as:

$$v_n = \Big[\underbrace{\sqrt{\tfrac{n}{\Delta_n^3}}, \dots, \sqrt{\tfrac{n}{\Delta_n^3}}}_{N_{\beta_{S_1}}}, \underbrace{\sqrt{\tfrac{n}{\Delta_n}}, \dots, \sqrt{\tfrac{n}{\Delta_n}}}_{N_{\beta_{S_2}}}, \underbrace{\sqrt{n\Delta_n}, \dots, \sqrt{n\Delta_n}}_{N_{\beta_R}}, \underbrace{\sqrt{n}, \dots, \sqrt{n}}_{N_\sigma}\Big]^\top.$$

The asymptotic normality immediately holds from the following two results — their proofs are shown in Appendices D.7 and D.8.

Lemma 13. *Assume that Condition (H)-II and (C1)–(C5) hold. If* $n \to \infty$, $\Delta_n \to 0$ *and* $n\Delta_n \to \infty$, *then*

$$\mathscr{I}_n\big(\theta^\dagger + \lambda(\hat{\theta}_n - \theta^\dagger)\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 2\Gamma(\theta^\dagger),$$

uniformly in $\lambda \in [0, 1]$, *where the matrix* $\Gamma(\theta^\dagger)$ *is defined as in (15) in the main text.*

Lemma 14. *Assume that Condition (H)-II and (C1)–(C5) hold. If* $n \to \infty$, $\Delta_n \to 0$ *and* $n\Delta_n \to \infty$, *with* $\Delta_n = o(n^{-1/2})$, *then*

$$\mathscr{C}_n(\theta^\dagger) \xrightarrow{\mathcal{L}_{\theta^\dagger}} \mathscr{N}\big(\mathbf{0}_{N_\theta}, 4\Gamma(\theta^\dagger)\big).$$

The proof of Theorem 2 is now complete.

Appendix D. Proof of technical results

D.1. Proof of Lemma 7

We have that $\frac{\Delta_n^3}{n}\ell_n(\theta) = \sum_{1\le i\le 4}\mathscr{E}_i(\theta)$, $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$, where we have set:

$$\mathscr{E}_1(\theta) = \frac{1}{n}\sum_{i=1}^n \eta_{S_1}(X_{i-1}, \beta_{S_1})^\top \Lambda_{S_1S_1}(X_{i-1}, \theta)\, \eta_{S_1}(X_{i-1}, \beta_{S_1});$$

$$\mathscr{E}_2(\theta) = \frac{1}{n}\sum_{i=1}^n \sum_{1\le j_1, j_2\le N} R_1^{j_1j_2}(\Delta_n^{q_1}, X_{i-1}, \theta)\, m_i^{j_1}(\Delta, \theta^\dagger)\, m_i^{j_2}(\Delta, \theta^\dagger);$$

$$\mathscr{E}_3(\theta) = \frac{1}{n}\sum_{i=1}^n \sum_{1\le j\le N} R_2^{j}(\Delta_n^{q_2}, X_{i-1}, \theta)\, m_i^{j}(\Delta, \theta^\dagger);$$

$$\mathscr{E}_4(\theta) = \frac{1}{n}\sum_{i=1}^n R_3(\Delta_n^{q_3}, X_{i-1}, \theta),$$

for some functions $R_1^{j_1j_2}, R_2^j, R_3 \in \mathcal{S}$ and constants $q_1, q_2, q_3 \ge 1$. From Lemmas 2, 3, we immediately have that as $n\to\infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$,

$$\mathscr{E}_1(\theta) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \eta_{S_1}(x, \beta_{S_1})^\top \Lambda_{S_1S_1}(x, \theta)\, \eta_{S_1}(x, \beta_{S_1}) \nu_{\theta^\dagger}(dx), \qquad \mathscr{E}_k(\theta) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \quad 2\le k\le 4,$$

uniformly in $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$, and the proof is now complete.

D.2. Proof of Lemma 8

We define $\mathscr{F} : \Theta \to \mathbb{R}^{N_{\beta_{S_1}}\times N_{\beta_{S_1}}}$ as:

$$\mathscr{F}(\theta) = \frac{\Delta_n^3}{n}\partial^2_{\beta_{S_1}} \ell_n(\theta), \quad \theta\in\Theta.$$

$\mathscr{F}(\theta)$ can be expressed as $\mathscr{F}(\theta) = \sum_{1\le k\le 6}\mathscr{F}_k(\theta)$, where we have set, for $1\le j_1, j_2 \le N_{\beta_{S_1}}$ with multi-index $\mathbf{j} = (j_1, j_2)$,

$$[\mathscr{F}_1(\theta)]_{j_1j_2} = \frac{2}{n}\sum_{i=1}^n \big(\partial_{j_1}^{\beta_{S_1}} V_{S_1,0}(X_{S,i-1}, \beta_{S_1})\big)^\top \Lambda_{S_1S_1}(X_{i-1},\theta)\partial_{j_2}^{\beta_{S_1}} V_{S_1,0}(X_{S,i-1}, \beta_{S_1});$$

$$[\mathscr{F}_2(\theta)]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^n \eta_{S_1}(X_{i-1}, \beta_{S_1})^\top \partial_{\mathbf{j}}^{\beta_{S_1}} \Lambda_{S_1S_1}(X_{i-1},\theta)\, \eta_{S_1}(X_{i-1}, \beta_{S_1});$$

$$[\mathscr{F}_3(\theta)]_{j_1j_2} = -\frac{2}{n}\sum_{i=1}^n \eta_{S_1}(X_{i-1}, \beta_{S_1})^\top \partial_{\mathbf{j}}^{\beta_{S_1}} \big\{\Lambda_{S_1S_1}(X_{i-1},\theta) V_{S_1,0}(X_{S,i-1}, \beta_{S_1})\big\};$$

$$[\mathscr{F}_4(\theta)]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^n \sum_{1\le k_1,k_2\le N} R_{k_1k_2}^{j_1j_2}(\Delta_n^{q_1}, X_{i-1}, \theta)\, m_i^{k_1}(\Delta_n, \theta^\dagger)\, m_i^{k_2}(\Delta_n, \theta^\dagger);$$

$$[\mathscr{F}_5(\theta)]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^n \sum_{1\le k\le N} R_{k}^{j_1j_2}(\Delta_n^{q_2}, X_{i-1}, \theta)\, m_i^{k}(\Delta_n, \theta^\dagger);$$

$$[\mathscr{F}_6(\theta)]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^n R^{j_1j_2}(\Delta_n^{q_3}, X_{i-1}, \theta),$$

for some functions $R_{k_1k_2}^{j_1j_2}, R_k^{j_1j_2}, R^{j_1j_2} \in \mathcal{S}$ and constants $q_1, q_2, q_3 > 0$. It follows from Lemmas 2, 3 and the consistency of estimator $\hat\beta_{S_1,n}$ that if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$,

$$\Big[\mathscr{F}_1\big(\beta_{S_1}^\dagger - \lambda(\hat\beta_{S_1,n} - \beta_{S_1}^\dagger), \beta_{S_2}, \beta_R, \sigma\big)\Big]_{j_1j_2}$$
$$\xrightarrow{\mathbb{P}_{\theta^\dagger}} 2\int \big(\partial_{j_1}^{\beta_{S_1}} V_{S_1,0}(x_S, \beta_{S_1}^\dagger)\big)^\top \Lambda_{S_1S_1}\big(x, (\beta_{S_1}^\dagger, \beta_{S_2}, \sigma)\big)\partial_{j_2}^{\beta_{S_1}} V_{S_1,0}(x_S, \beta_{S_1}^\dagger)\nu_{\theta^\dagger}(dx);$$

$$\Big[\mathscr{F}_k\big(\beta_{S_1}^\dagger - \lambda(\hat\beta_{S_1,n} - \beta_{S_1}^\dagger), \beta_{S_2}, \beta_R, \sigma\big)\Big]_{j_1j_2} \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \qquad 2\le k\le 6,$$

uniformly in $(\beta_{S_2}, \beta_R, \sigma) \in \Theta_{\beta_{S_2}}\times\Theta_{\beta_R}\times\Theta_\sigma$ and $\lambda\in[0,1]$. The proof is now complete.

D.3. Proof of Lemma 9

We write $\theta^{-S_1} \equiv \left(\beta_{S_2}, \beta_R, \sigma\right) \in \Theta_{\beta_{S_2}} \times \Theta_{\beta_R} \times \Theta_\sigma$. It holds that

$$\tfrac{\Delta_n}{n}\,\ell_n(\hat{\beta}_{S_1,n},\,\theta^{-S_1}) = \sum_{1\le k\le 7} \mathscr{G}_k(\hat{\beta}_{S_1,n},\,\theta^{-S_1}),$$

where we have set, for $\theta \in \Theta$,

$$\mathscr{G}_1(\theta) = \tfrac{\Delta_n}{n}\sum_{i=1}^n \eta_{S_1,i-1}(\sqrt{\Delta_n^3},\beta_{S_1})^\top \Lambda_{S_1S_1}(X_{i-1},\theta)\eta_{S_1,i-1}(\sqrt{\Delta_n^3},\beta_{S_1});$$

$$\mathscr{G}_2(\theta) = \tfrac{1}{n}\sum_{i=1}^n \sum_{1\le j\le N} R_j(\sqrt{\Delta_n},X_{i-1},\theta)\eta^j_{S_1,i-1}(\sqrt{\Delta_n^3},\beta_{S_1});$$

$$\mathscr{G}_3(\theta) = \tfrac{1}{n}\sum_{i=1}^n \sum_{1\le j_1,j_2\le N} R_{j_1j_2}(\Delta_n,X_{i-1},\theta)\eta^{j_1}_{S_1,i-1}(\sqrt{\Delta_n^3},\beta_{S_1})m_i^{j_2}(\Delta_n,\theta^\dagger);$$

$$\begin{aligned}\mathscr{G}_4(\theta) = \tfrac{1}{n}\sum_{i=1}^n \Bigg\{ &\tfrac14\Big(\mathcal{L}V_{S_1,0}(X_{i-1},\theta^\dagger)-\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\Big)^\top \Lambda_{S_1S_1}(X_{i-1},\theta)\Big(\mathcal{L}V_{S_1,0}(X_{i-1},\theta^\dagger)-\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\Big)\\
&+\tfrac12\Big(\mathcal{L}V_{S_1,0}(X_{i-1},\theta^\dagger)-\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\Big)^\top \Lambda_{S_1S_2}(X_{i-1},\theta)\,\eta_{S_2}(X_{i-1},\beta_{S_2})\\
&+\tfrac12\eta_{S_2}(X_{i-1},\beta_{S_2})^\top\Lambda_{S_2S_1}(X_{i-1},\theta)\Big(\mathcal{L}V_{S_1,0}(X_{i-1},\theta^\dagger)-\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\Big)\\
&+\eta_{S_2}(X_{i-1},\beta_{S_2})^\top\Lambda_{S_2S_2}(X_{i-1},\theta)\eta_{S_2}(X_{i-1},\beta_{S_2})\Bigg\};\end{aligned}$$

$$\mathscr{G}_5(\theta) = \tfrac{1}{n}\sum_{i=1}^n \sum_{1\le j\le N} \widetilde{R}_j(\Delta_n,X_{i-1},\theta)\,m_i^j(\Delta_n,\theta^\dagger);$$

$$\mathscr{G}_6(\theta) = \tfrac{\Delta_n}{n}\sum_{i=1}^n \Big\{ m_i(\Delta_n,\theta^\dagger)^\top\Lambda(X_{i-1},\theta)\,m_i(\Delta_n,\theta^\dagger)+\log\Big|\Sigma(X_{i-1},\theta)\Big|\Big\},$$

for some functions $R_j, R_{j_1j_2}, \widetilde{R}_j \in \mathcal{S}$. Note that $\eta^j_{S_1,i-1}(\sqrt{\Delta_n^3},\hat{\beta}_{S_1,n})$ can be expressed as:

$$\eta^j_{S_1,i-1}(\sqrt{\Delta_n^3},\hat{\beta}_{S_1,n}) = \sum_{1\le k\le N_{\beta_{S_1}}} \frac{\eta^{j,[k]}_{S_1}(X_{S,i-1})}{|\beta^{\dagger,k}_{S_1}-\hat{\beta}^k_{S_1,n}|}\times\left|\frac{\beta^{\dagger,k}_{S_1}-\hat{\beta}^k_{S_1,n}}{\sqrt{\Delta_n^3}}\right|,$$

where we have set:

$$\eta^{j,[k]}_{S_1}(X_{S,i-1}) \equiv V^j_{S_1,0}(X_{S,i-1},\bar{\beta}^{[k-1]}_{S_1,n}) - V^j_{S_1,0}(X_{S,i-1},\bar{\beta}^{[k]}_{S_1,n}),\quad 1\le k\le N_{\beta_{S_1}},$$

with the notation:

$$\bar{\beta}^{[\ell]}_{S_1,n} = (\hat{\beta}^1_{S_1,n},\dots,\hat{\beta}^\ell_{S_1,n},\beta^{\dagger,\ell+1}_{S_1},\dots,\beta^{\dagger,N_{\beta_{S_1}}}_{S_1}),\quad 1\le\ell\le N_{\beta_{S_1}}-1;$$

$$\bar{\beta}^{[0]}_{S_1,n} = \beta^\dagger_{S_1},\quad \bar{\beta}^{[N_{\beta_{S_1}}]}_{S_1,n} = \hat{\beta}_{S_1,n}.$$

From convergence (43), Lemma 2 and condition (C2) it follows that if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then

$$\tfrac1n\sum_{i=1}^n f(X_{i-1},\theta)\eta^j_{S_1,i-1}(\sqrt{\Delta_n^3},\hat{\beta}_{S_1,n})\xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \tag{49}$$

uniformly in $\theta\in\Theta=(\beta_{S_2},\beta_R,\sigma)\in\Theta_{\beta_{S_2}}\times\Theta_{\beta_R}\times\Theta_\sigma$ for any $f:\mathbb{R}^N\times\Theta\to\mathbb{R}$ satisfying the same property in Lemma 2. Thus, we obtain

$$\mathscr{G}_k(\hat{\beta}_{S_1,n},\theta^{-S_1})\xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,\qquad 1\le k\le 3,$$

uniformly in θ^{-S_1}. Also, we immediately have from Lemmas 2, 3 that

$$\mathscr{G}_k(\hat{\beta}_{S_1,n},\theta^{-S_1})\xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,\qquad k=5,6,$$

as $n\to\infty$, $\Delta_n\to0$ and $n\Delta_n\to\infty$, uniformly in θ^{-S_1}. Finally, we consider the fourth term $\mathscr{G}_4$. Noticing that

$$\begin{aligned}&\mathcal{L}V_{S_1,0}(X_{i-1},\theta^\dagger)-\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\\
&\quad=\partial^\top_{x_{S_1}}V_{S_1,0}(X_{S,i-1},\beta^\dagger_{S_1})\,\eta_{S_1}(X_{i-1},\beta_{S_1})+\Big(\partial^\top_{x_{S_1}}V_{S_1,0}(X_{S,i-1},\beta^\dagger_{S_1})-\partial^\top_{x_{S_1}}V_{S_1,0}(X_{S,i-1},\beta_{S_1})\Big)V_{S_1,0}(X_{S,i-1},\beta_{S_1})\\
&\qquad+\partial^\top_{x_{S_2}}V_{S_1,0}(X_{S,i-1},\beta^\dagger_{S_1})\,\eta_{S_2}(X_{i-1},\beta_{S_2})+\Big(\partial^\top_{x_{S_2}}V_{S_1,0}(X_{S,i-1},\beta^\dagger_{S_1})-\partial^\top_{x_{S_2}}V_{S_1,0}(X_{S,i-1},\beta_{S_1})\Big)V_{S_2,0}(X_{i-1},\beta_{S_2}),\end{aligned}$$

we obtain from Lemmas 2, 3 and consistency of $\hat{\beta}_{S_1,n}$ that

$$\mathscr{G}_4(\hat{\beta}_{S_1,n},\theta^{-S_1}) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \eta_{S_2}(x,\beta_{S_2})^\top \widetilde{\mathscr{G}}_4\big(x,(\beta_{S_1}^\dagger,\beta_{S_2},\beta_R,\sigma)\big)\,\eta_{S_2}(x,\beta_{S_2})\,\nu_{\theta^\dagger}(dx),$$

as $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, uniformly in θ^{-S_1}, where we have set: for $x=(x_S,x_R)\in\mathbb{R}^{N_S}\times\mathbb{R}^{N_R}$ and $\theta=(\beta_{S_1},\beta_{S_2},\beta_R,\sigma)\in\Theta$,

$$\begin{aligned}\widetilde{\mathscr{G}}_4\big(x,\theta\big) &= \tfrac14\big(\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})\big)^\top \Lambda_{S_1S_1}(x,\theta)\,\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})\\ &\quad+\tfrac12\big(\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})\big)^\top\Lambda_{S_1S_2}(x,\theta)+\tfrac12\Lambda_{S_2S_1}(x,\theta)\,\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})+\Lambda_{S_2S_2}(x,\theta).\end{aligned}$$

Lemmas 4, 5 yield:

$$\widetilde{\mathscr{G}}_4\big(x,\theta\big)=\tfrac12\Lambda_{S_2S_1}(x,\theta)\,\partial_{x_{S_2}}^\top V_{S_1,0}(x_S,\beta_{S_1})+\Lambda_{S_2S_2}(x,\theta)=12a_{S_2}^{-1}(x,\theta).$$

The proof is now complete.

D.4. Proof of Lemma 10

(*Proof of* (47)). Since the proof of $\frac{\sqrt{\Delta_n^3}}{n}\partial_{\beta_{S_1}}\ell_n(\beta_S^\dagger,\beta_R,\sigma)\xrightarrow{\mathbb{P}_{\theta^\dagger}}\mathbf{0}_{N_{\beta_{S_1}}}$ is identical with that in Appendix C.1.1, we will only show that, if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then

$$\mathscr{K}(\beta_S^\dagger,\beta_R,\sigma) := \frac{\sqrt{\Delta_n}}{n}\partial_{\beta_{S_2}}\ell_n(\beta_S^\dagger,\beta_R,\sigma)\xrightarrow{\mathbb{P}_{\theta^\dagger}}\mathbf{0}_{N_{\beta_{S_2}}}, \tag{50}$$

uniformly in $(\beta_R,\sigma)\in\Theta_{\beta_R}\times\Theta_\sigma$. It holds that $\mathscr{K}(\beta_S^\dagger,\beta_R,\sigma)=\sum_{1\le l\le 3}\mathscr{K}_l(\beta_S^\dagger,\beta_R,\sigma)$, where we have set, for $\theta\in\Theta$,

$$\begin{aligned}\mathscr{K}_1(\theta) &= \frac1n\sum_{i=1}^n\sum_{1\le j\le N}R_j(1,X_{i-1},\theta)\,m_i^j(\Delta_n,\theta^\dagger);\\ \mathscr{K}_2(\theta) &= \frac1n\sum_{i=1}^n\sum_{1\le j_1,j_2\le N}R_{j_1j_2}(\sqrt{\Delta_n},X_{i-1},\theta)\,m_i^{j_1}(\Delta_n,\theta^\dagger)\,m_i^{j_2}(\Delta_n,\theta^\dagger);\\ \mathscr{K}_3(\theta) &= \frac1n\sum_{i=1}^n R(\sqrt{\Delta_n},X_{i-1},\theta),\end{aligned}$$

for $R,R_j,R_{j_1j_2}\in\mathcal{S}$. From Lemmas 2 and 3, we immediately have that $\mathscr{K}_i(\beta_S^\dagger,\beta_R,\sigma)\xrightarrow{\mathbb{P}_{\theta^\dagger}}\mathbf{0}_{N_{\beta_{S_2}}}$, $1\le i\le 3$, as $n\to\infty$, $\Delta_n\to0$ and $n\Delta_n\to\infty$, thus (50) holds.

(*Proof of* (48)). We define

$$Q(\theta)=M_{\beta_S,n}\,\partial_{\beta_S}^2\ell_n(\theta)\,M_{\beta_S,n}=\begin{bmatrix}Q_{S_1S_1}(\theta) & Q_{S_1S_2}(\theta)\\ Q_{S_2S_1}(\theta) & Q_{S_2S_2}(\theta)\end{bmatrix},$$

where we have set:

$$\begin{aligned}&Q_{S_1S_1}(\theta)=\tfrac{\Delta_n^3}{n}\partial_{\beta_{S_1}}^2\ell_n(\theta),\quad Q_{S_1S_2}(\theta)=\tfrac{\Delta_n^2}{n}\partial_{\beta_{S_1}}\partial_{\beta_{S_2}}^\top\ell_n(\theta);\\ &Q_{S_2S_1}(\theta)=Q_{S_1S_2}(\theta)^\top,\quad Q_{S_2S_2}(\theta)=\tfrac{\Delta_n}{n}\partial_{\beta_{S_2}}^2\ell_n(\theta),\end{aligned}$$

for $\theta=(\beta_S,\beta_R,\sigma)\in\Theta$. From the proof of Lemma 8 in Appendix D.2 and the consistency of the estimator $\hat{\beta}_{S,n}$, we have that if $n\to\infty$, $\Delta_n\to0$ and $n\Delta_n\to\infty$, then

$$Q_{S_1S_1}\big(\beta_S^\dagger+\lambda(\hat{\beta}_{S,n}-\beta_S^\dagger),\beta_R,\sigma\big)\xrightarrow{\mathbb{P}_{\theta^\dagger}}2\int\partial_{\beta_{S_1}}\big(V_{S_1,0}(x_S,\beta_{S_1}^\dagger)\big)^\top\Lambda_{S_1S_1}\big(x,(\beta_S^\dagger,\sigma)\big)\partial_{\beta_{S_1}}^\top V_{S_1,0}(x_S,\beta_{S_1}^\dagger)\,\nu_{\theta^\dagger}(dx), \tag{51}$$

uniformly in $(\beta_R,\sigma)\in\Theta_{\beta_R}\times\Theta_\sigma$ and $\lambda\in[0,1]$. We will check the convergence of the two matrices $Q_{S_1S_2}(\theta)$ and $Q_{S_2S_2}(\theta)$.

We have $Q_{S_1S_2}(\theta)=\sum_{1\le k\le4}Q_{S_1S_2,k}(\theta)$, where we set for $1\le j_1\le N_{\beta_{S_1}}$, $1\le j_2\le N_{\beta_{S_2}}$,

$$\begin{aligned}\big[Q_{S_1S_2,1}(\theta)\big]_{j_1j_2} &= \frac2n\sum_{i=1}^n\big(\partial_{j_1}^{\beta_{S_1}}V_{S_1,0}(X_{S,i-1},\beta_{S_1})\big)^\top\times\Big\{\Lambda_{S_1S_2}\big(X_{i-1},\theta\big)\partial_{j_2}^{\beta_{S_2}}V_{S_2,0}(X_{i-1},\beta_{S_2})\\ &\qquad+\tfrac12\Lambda_{S_1S_1}\big(X_{i-1},\theta\big)\partial_{j_2}^{\beta_{S_2}}\mathcal{L}V_{S_1,0}(X_{i-1},\beta_S)\Big\};\\ \big[Q_{S_1S_2,2}(\theta)\big]_{j_1j_2} &= \frac1n\sum_{i=1}^n\sum_{1\le k_1,k_2\le N_{S_1}}R_{j_1j_2}^{k_1k_2}(1,X_{i-1},\theta)\eta_{S_1,i-1}^{k_1}(\Delta_n,\beta_{S_1})\eta_{S_1}^{k_2}(X_{i-1},\beta_{S_1});\\ \big[Q_{S_1S_2,3}(\theta)\big]_{j_1j_2} &= \frac1n\sum_{i=1}^n\sum_{\substack{1\le k_1\le N_{S_1}\\1\le k_2\le N_S}}\widetilde{R}_{j_1j_2}^{k_1k_2}(1,X_{i-1},\theta)\eta_{S_1}^{k_1}(X_{i-1},\beta_{S_1})b^{k_2}(X_{i-1},\beta_S);\end{aligned}$$

$$\big[Q_{S_1S_2,4}(\theta)\big]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^{n}\Big\{\sum_{1\le k_1,k_2\le N}\bar{R}^{k_1k_2}_{j_1j_2}(\Delta_n, X_{i-1},\theta)m_i^{k_1}(\Delta_n,\theta^\dagger)m_i^{k_2}(\Delta_n,\theta^\dagger) + \sum_{1\le k\le N}\bar{R}^{k}_{j_1j_2}(\Delta_n, X_{i-1},\theta)m_i^{k}(\Delta_n,\theta^\dagger) + \bar{R}(\Delta_n, X_{i-1},\theta)\Big\},$$

for some functions $R^{k_1k_2}_{j_1j_2}$, $\widetilde{R}^{k_1k_2}_{j_1j_2}$, $\bar{R}^{k_1k_2}_{j_1j_2}$, $\bar{R}^{k}_{j_1j_2}$, $\bar{R}\in\mathcal{S}$, and the function $b:\mathbb{R}^N\times\Theta_{\beta_S}\to\mathbb{R}^{N_S}$ is defined as:

$$b(x,\beta_S) = \begin{bmatrix}\frac{1}{2}\mathcal{L}V_{S_1,0}(x,\beta_S) - \frac{1}{2}\mathcal{L}V_{S_1,0}(x,\beta_S^\dagger)\\ V_{S_2,0}(x,\beta_{S_2}) - V_{S_2,0}(x,\beta_{S_2}^\dagger)\end{bmatrix},$$

Notice that for any $\theta\in\Theta$,

$$[Q_{S_1S_2,1}(\theta)]_{j_1j_2} = 0, \quad 1\le j_1\le N_{\beta_{S_1}}, \quad 1\le j_2\le N_{\beta_{S_2}},$$

because it follows from Lemma 4 that

$$\Lambda_{S_1S_2}(x,\theta)\partial^{\beta_{S_2}}_{j_2}V_{S_2,0}(x,\beta_{S_2}) + \frac{1}{2}\Lambda_{S_1S_1}(x,\theta)\partial^{\beta_{S_2}}_{j_2}\mathcal{L}V_{S_1,0}(x,\beta_S) = \Phi(x,\theta)\,\partial^{\beta_{S_2}}_{j_2}V_{S_2,0}(x,\beta_{S_2}) = \mathbf{0}_{N_{S_1}},$$

for any $x\in\mathbb{R}^N$, $1\le j_2\le N_{\beta_{S_2}}$, where $\Phi(x,\theta)$ is defined as in (46). From Lemmas 2, 3 and the consistency of $\hat{\beta}_{S,n}$ with convergence (43), we obtain:

$$\big[Q_{S_1S_2,k}\big(\beta_S^\dagger + \lambda(\hat{\beta}_{S,n}-\beta_S^\dagger),\beta_R,\sigma\big)\big]_{j_1j_2} \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \qquad 2\le k\le 4,$$

uniformly in $(\beta_R,\sigma)\in\Theta_{\beta_R}\times\Theta_\sigma$ and $\lambda\in[0,1]$.

Finally, we consider the term $Q_{S_2S_2}(\theta)$. It holds that $Q_{S_2S_2}(\theta)=\sum_{1\le k\le 4}Q_{S_2S_2,k}(\theta)$, where we set, for $1\le j_1,j_2\le N_{\beta_{S_2}}$,

$$\begin{aligned}\big[Q_{S_2S_2,1}(\theta)\big]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\sum_{1\le k\le N_{S_1}}R^{k}_{j_1j_2}(1,X_{i-1},\theta)\eta^{k}_{S_1,i-1}(\Delta_n,\beta_{S_1})\\ &+\sum_{1\le k_1,k_2\le N_{S_1}}R^{k_1k_2}_{j_1j_2}(1,X_{i-1},\theta)\eta^{k_1}_{S_1,i-1}(\Delta_n,\beta_{S_1})\eta^{k_2}_{S_1,i-1}(\Delta_n,\beta_{S_1})\\ &+\sum_{\substack{1\le k_1\le N_{S_1}\\1\le k_2\le N}}\widetilde{R}^{k_1k_2}_{j_1j_2}(1,X_{i-1},\theta)\eta^{k_1}_{S_1,i-1}(\sqrt{\Delta_n},\beta_{S_1})m_i^{k_2}(\Delta_n,\theta^\dagger)\Big\};\end{aligned}$$

$$\begin{aligned}\big[Q_{S_2S_2,2}(\theta)\big]_{j_1j_2} = \frac{2}{n}\sum_{i=1}^{n}\Big\{&\frac{1}{4}\big(\partial^{\beta_{S_2}}_{j_1}\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\big)^\top\Lambda_{S_1S_1}(X_{i-1},\theta)\partial^{\beta_{S_2}}_{j_2}\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\\ &+\frac{1}{2}\big(\partial^{\beta_{S_2}}_{j_1}\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\big)^\top\Lambda_{S_1S_2}(X_{i-1},\theta)\,\partial^{\beta_{S_2}}_{j_2}V_{S_2,0}(X_{i-1},\beta_{S_2})\\ &+\frac{1}{2}\big(\partial^{\beta_{S_2}}_{j_1}V_{S_2,0}(X_{i-1},\beta_{S_2})\big)^\top\Lambda_{S_2S_1}(X_{i-1},\theta)\partial^{\beta_{S_2}}_{j_2}\mathcal{L}V_{S_1,0}(X_{i-1},\theta)\\ &+\big(\partial^{\beta_{S_2}}_{j_1}V_{S_2,0}(X_{i-1},\beta_{S_2})\big)^\top\Lambda_{S_2S_2}(X_{i-1},\theta)\,\partial^{\beta_{S_2}}_{j_2}V_{S_2,0}(X_{i-1},\beta_{S_2})\Big\};\end{aligned}$$

$$\begin{aligned}\big[Q_{S_2S_2,3}(\theta)\big]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\sum_{1\le k_1,k_2\le N_{S_2}}\overline{R}^{k_1k_2}_{j_1j_2}(1,X_{i-1},\theta)\eta^{k_1}_{S_2}(X_{i-1},\beta_{S_2})\eta^{k_2}_{S_2}(X_{i-1},\beta_{S_2})\\ &+\sum_{1\le k\le N_{S_2}}\widetilde{R}^{k}_{j_1j_2}(1,X_{i-1},\theta)\eta^{k}_{S_2}(X_{i-1},\beta_{S_2})\Big\};\end{aligned}$$

$$\begin{aligned}\big[Q_{S_2S_2,4}(\theta)\big]_{j_1j_2} = \frac{1}{n}\sum_{i=1}^{n}\Big\{&\sum_{1\le k_1,k_2\le N}\hat{R}^{k_1k_2}_{j_1j_2}(\Delta_n,X_{i-1},\theta)m_i^{k_1}(\Delta_n,\theta^\dagger)m_i^{k_2}(\Delta_n,\theta^\dagger)\\ &+\sum_{1\le k\le N}\overline{R}^{k}_{j_1j_2}(\sqrt{\Delta_n},X_{i-1},\theta)m_i^{k}(\Delta_n,\theta^\dagger) + R_{j_1j_2}(\sqrt{\Delta_n},X_{i-1},\theta)\Big\},\end{aligned}$$

for some functions $R^{k}_{j_1j_2}$, $R^{k_1k_2}_{j_1j_2}$, $\widetilde{R}^{k_1k_2}_{j_1j_2}$, $\overline{R}^{k_1k_2}_{j_1j_2}$, $\widetilde{R}^{k}_{j_1j_2}$, $\hat{R}^{k_1k_2}_{j_1j_2}$, $\overline{R}^{k}_{j_1j_2}$, $R_{j_1j_2}\in\mathcal{S}$. Note that due to Lemma 4,

$$[Q_{S_2S_2,2}(\theta)]_{j_1j_2} = \frac{2}{n}\sum_{i=1}^{n}\frac{1}{2}\big(\partial^{\beta_{S_2}}_{j_1}V_{S_2,0}(X_{i-1},\beta_{S_2})\big)^\top\Lambda_{S_2S_1}(X_{i-1},\theta)\,\partial^{\beta_{S_2}}_{j_2}\mathcal{L}V_{S_1,0}(X_{i-1},\theta)$$
$$+\frac{2}{n}\sum_{i=1}^{n}\big(\partial^{\beta_{S_2}}_{j_1}V_{S_2,0}(X_{i-1},\beta_{S_2})\big)^\top\Lambda_{S_2S_2}(X_{i-1},\theta)\,\partial^{\beta_{S_2}}_{j_2}V_{S_2,0}(X_{i-1},\beta_{S_2}).$$

We obtain from Lemma 2, (49) and consistency of $\hat{\beta}_{S,n}$ that if $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, then

$$\Big[Q_{S_2S_2,2}\big(\beta_S^\dagger + \lambda(\hat{\beta}_{S,n}-\beta_S^\dagger),\beta_R,\sigma\big)\Big]_{j_1j_2}$$

$$
\begin{aligned}
&\xrightarrow{\mathbb{P}_{\theta^\dagger}} 2\int \tfrac{1}{2}\left(\partial_{j_1}^{\beta_{S_2}} V_{S_2,0}(x,\beta_{S_2}^\dagger)\right)^\top \Lambda_{S_2S_1}\left(x,(\beta_S^\dagger,\beta_R,\sigma)\right)\partial_{j_2}^{\beta_{S_2}}\mathcal{L}V_{S_1,0}(x,(\beta_S^\dagger,\beta_R,\sigma))\nu_{\theta^\dagger}(dx)\\
&\qquad +2\int \left(\partial_{j_1}^{\beta_{S_2}} V_{S_2,0}(x,\beta_{S_2}^\dagger)\right)^\top \Lambda_{S_2S_2}\left(x,(\beta_S^\dagger,\beta_R,\sigma)\right)\partial_{j_2}^{\beta_{S_2}} V_{S_2,0}(x,\beta_{S_2}^\dagger)\nu_{\theta^\dagger}(dx); \qquad (52)\\
&\left[Q_{S_2S_2,k}\left(\beta_S^\dagger+\lambda(\hat\beta_{S,n}-\beta_S^\dagger),\beta_R,\sigma\right)\right]_{j_1j_2}\xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \qquad k=1,3,4.
\end{aligned}
$$

The proof is complete by applying Lemma 5 to (51) and (52).

D.5. Proof of Lemma 11

It holds that $\frac{1}{n}\ell_n(\theta)=\sum_{1\le k\le 4}\mathscr{T}_k(\theta)$, $\theta\in\Theta$, where we have set:

$$
\begin{aligned}
\mathscr{T}_1(\theta) &= \frac{1}{n}\sum_{i=1}^n\Bigg\{\sum_{1\le j_1,j_2\le N_{S_1}} R_{j_1j_2}^{S_1S_1}(1,X_{i-1},\theta)\eta_{S_1,i-1}^{j_1}(\sqrt{\Delta_n^3},\beta_{S_1})\eta_{S_1,i-1}^{j_2}(\sqrt{\Delta_n^3},\beta_{S_1})\\
&\quad+\sum_{\substack{1\le j_1\le N_{S_1}\\1\le j_2\le N_{S_2}}} R_{j_1j_2}^{S_1S_2}(1,X_{i-1},\theta)\eta_{S_1,i-1}^{j_1}(\sqrt{\Delta_n^3},\beta_{S_1})\eta_{S_2,i-1}^{j_2}(\sqrt{\Delta_n},\beta_{S_2})\\
&\quad+\sum_{1\le j_1,j_2\le N_{S_2}} R_{j_1j_2}^{S_2S_2}(1,X_{i-1},\theta)\eta_{S_2,i-1}^{j_1}(\sqrt{\Delta_n},\beta_{S_2})\eta_{S_2,i-1}^{j_2}(\sqrt{\Delta_n},\beta_{S_2})\Bigg\};\\
\mathscr{T}_2(\theta) &= \frac{1}{n}\sum_{i=1}^n\Bigg\{\sum_{\substack{1\le j_1\le N_{S_1}\\1\le j_2\le N}} R_{j_1j_2}^{S_1}(1,X_{i-1},\theta)\eta_{S_1,i-1}^{j_1}(\sqrt{\Delta_n^3},\beta_{S_1})m_i^{j_2}(\Delta_n,\theta^\dagger)\\
&\quad+\sum_{\substack{1\le j_1\le N_{S_2}\\1\le j_2\le N}} R_{j_1j_2}^{S_2}(1,X_{i-1},\theta)\eta_{S_2,i-1}^{j_1}(\sqrt{\Delta_n},\beta_{S_2})m_i^{j_2}(\Delta_n,\theta^\dagger)\\
&\quad+\sum_{1\le j\le N_{S_1}} R_j^{S_1}(\sqrt{\Delta_n},X_{i-1},\theta)\eta_{S_1,i-1}^{j}(\sqrt{\Delta_n^3},\beta_{S_1})+\sum_{1\le j\le N_{S_2}} R_j^{S_2}(\sqrt{\Delta_n},X_{i-1},\theta)\eta_{S_2,i-1}^{j}(\sqrt{\Delta_n},\beta_{S_2})\Bigg\};\\
\mathscr{T}_3(\theta) &= \frac{1}{n}\sum_{i=1}^n\Bigg\{m_i(\Delta_n,\theta^\dagger)^\top\Lambda\left(X_{i-1},(\beta_S,\sigma)\right)m_i(\Delta_n,\theta^\dagger)+\log\left|\Sigma\left(X_{i-1},(\beta_S,\sigma)\right)\right|\Bigg\};\\
\mathscr{T}_4(\theta) &= \frac{1}{n}\sum_{i=1}^n\Bigg\{\sum_{1\le j\le N} R_j(\sqrt{\Delta_n},X_{i-1},\theta)m_i^j(\Delta_n,\theta^\dagger)+R(\Delta_n,X_{i-1},\theta)\Bigg\},
\end{aligned}
$$

where $R_{j_1j_2}^{S_1S_1}$, $R_{j_1j_2}^{S_1S_2}$, $R_{j_1j_2}^{S_2S_2}$, $R_{j_1j_2}^{S_1}$, $R_{j_1j_2}^{S_2}$, $R_j^{S_1}$, $R_j^{S_2}$, $R_{j_1j_2}$ R_j, $R\in\mathcal{S}$. From Lemmas 2, 3 and the convergences (43) and (44), we have:

$$
\begin{aligned}
&\mathscr{T}_k\left(\hat\beta_{S,n},\beta_R,\sigma\right)\xrightarrow{\mathbb{P}_{\theta^\dagger}} 0, \qquad k=1,2,4;\\
&\mathscr{T}_3\left(\hat\beta_{S,n},\beta_R,\sigma\right)\xrightarrow{\mathbb{P}_{\theta^\dagger}} \int\left\{\mathrm{tr}\left(\Lambda\left(x,(\beta_S^\dagger,\sigma)\right)\Sigma\left(x,(\beta_S^\dagger,\sigma^\dagger)\right)\right)+\log\left|\Sigma\left(x,(\beta_S^\dagger,\sigma)\right)\right|\right\}\nu_{\theta^\dagger}(dx),
\end{aligned}
$$

as $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, uniformly in $(\beta_R,\sigma)\in\Theta_{\beta_R}\times\Theta_\sigma$. The proof is now complete.

D.6. Proof of Lemma 12

The matrix-valued function $\Sigma(x,\theta)$ and its inverse $\Lambda(x,\theta)$ depend on $\beta_S=(\beta_{S_1},\beta_{S_2})\in\Theta_{\beta_S}$ and $\sigma\in\Theta_\sigma$ but not on $\beta_R\in\Theta_{\beta_R}$ in terms of the parameter $\theta\in\Theta$. We then define

$$
\mathscr{L}(\theta)=\frac{1}{n\Delta_n}\ell_n(\theta)-\frac{1}{n\Delta_n}\ell_n\left(\beta_S,\beta_R^\dagger,\sigma\right), \quad \theta=(\beta_S,\beta_R,\sigma)\in\Theta,
$$

and $\mathscr{L}(\theta)$ is expressed as $\mathscr{L}(\theta)=\sum_{1\le k\le 3}\mathscr{U}_k(\theta)$, where we have set:

$$
\begin{aligned}
\mathscr{U}_1(\theta) &= \frac{1}{n\Delta_n}\sum_{i=1}^n\left\{\left(m_i(\Delta_n,(\beta_S,\beta_R^\dagger,\sigma))-m_i(\Delta_n,\theta^\dagger)\right)^\top\Lambda(X_{i-1},(\beta_S,\sigma))\left(m_i(\Delta_n,\theta)-m_i(\Delta_n,(\beta_S,\beta_R^\dagger,\sigma))\right)\right\};\\
\mathscr{U}_2(\theta) &= \frac{1}{n\Delta_n}\sum_{i=1}^n\left\{\left(m_i(\Delta_n,\theta)-m_i(\Delta_n,\theta^\dagger)\right)^\top\Lambda(X_{i-1},(\beta_S,\sigma))\left(m_i(\Delta_n,\theta)-m_i(\Delta_n,(\beta_S,\beta_R^\dagger,\sigma))\right)\right\};\\
\mathscr{U}_3(\theta) &= \frac{2}{n\Delta_n}\sum_{i=1}^n m_i(\Delta_n,\theta^\dagger)^\top\Lambda(X_{i-1},(\beta_S,\sigma))\left(m_i(\Delta_n,\theta)-m_i(\Delta_n,(\beta_S,\beta_R^\dagger,\sigma))\right).
\end{aligned}
$$

We will derive the limit of the terms $\mathscr{U}_k(\theta)$, $1 \le k \le 3$ evaluated at $\theta = (\hat{\beta}_{S,n}, \beta_R, \hat{\sigma}_n)$ by utilising Lemma 6: We first consider the term $\mathscr{U}_1(\theta)$. Making use of Lemma 6, we have

$$\begin{aligned}\mathscr{U}_1(\theta) = \frac{1}{n}\sum_{i=1}^{n}\Bigg\{ &\sum_{1\le j\le N_{S_1}} \eta^j_{S_1,i-1}(\sqrt{\Delta_n^3}, \beta_{S_1}) R^j_{S_1}(\sqrt{\Delta_n}, X_{i-1}, \theta) \\ &+ \sum_{1\le j\le N_{S_2}} \eta^j_{S_2,i-1}(\sqrt{\Delta_n}, \beta_{S_2}) R^j_{S_2}(\sqrt{\Delta_n}, X_{i-1}, \theta) + R(\sqrt{\Delta_n}, X_{i-1}, \theta) \Bigg\},\end{aligned} \tag{53}$$

where $R^j_{S_1}$, $R^j_{S_2}$, $R \in \mathcal{S}$. From Lemma 2 and the limits (43), (44), we obtain that if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$,

$$\mathscr{U}_1\big(\hat{\beta}_{S,n}, \beta_R, \hat{\sigma}_n\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

uniformly in $\beta_R \in \Theta_{\beta_R}$. For the term $\mathscr{U}_2(\theta)$, again Lemma 6 yields

$$\mathscr{U}_2(\theta) = \frac{1}{n}\sum_{i=1}^{n} \eta_R(X_{i-1}, \beta_R)^\top a_R^{-1}(X_{i-1}, \sigma)\eta_R(X_{i-1}, \beta_R) + \widetilde{\mathscr{U}_2}(\theta),$$

where $\widetilde{\mathscr{U}_2}(\theta)$ is given in the form of the right-hand side of formula (53). We then obtain

$$\mathscr{U}_2\big(\hat{\beta}_{S,n}, \beta_R, \hat{\sigma}_n\big) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \eta_R(x, \beta_R)^\top\, a_R^{-1}(x, \sigma^\dagger)\, \eta_R(x, \beta_R) \nu_{\theta^\dagger}(dx),$$

as $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, uniformly in $\beta_R \in \Theta_{\beta_R}$. For the third term $\mathscr{U}_3(\theta)$, it follows from Lemma 6 that

$$\mathscr{U}_3(\theta) = \frac{1}{n\sqrt{\Delta_n}}\sum_{i=1}^{n}\sum_{N_S+1\le j\le N} m_i^j(\Delta_n, \theta^\dagger) R^j(1, X_{i-1}, \theta) + \frac{1}{n}\sum_{i=1}^{n}\sum_{1\le j\le N} m_i^j(\Delta_n, \theta^\dagger)\widetilde{R}^j(\sqrt{\Delta_n}, X_{i-1}, \theta),$$

where R^j, $\widetilde{R}^j \in \mathcal{S}$. From Lemma 3, we have that, if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$\mathscr{U}_3(\hat{\beta}_{S,n}, \beta_R, \hat{\sigma}_n) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

uniformly in $\beta_R \in \Theta_{\beta_R}$. The proof is now complete.

D.7. Proof of Lemma 13

Recall $N_\theta = N_\beta + N_\sigma$, where $N_\beta = N_{\beta_S} + N_{\beta_R}$ with $N_{\beta_S} = N_{\beta_{S_1}} + N_{\beta_{S_2}}$. In this section we make use of the notation $\beta = (\beta_S, \beta_R) \in \Theta_{\beta_S} \times \Theta_{\beta_R}$ with $\beta_S = (\beta_{S_1}, \beta_{S_2})$. We note again that the matrices $\Sigma(x,\theta)$ and $\Lambda(x,\theta)$ do not depend on the parameter β_R. Since we have seen the convergences of the matrices $Q_{S_1S_1}(\theta)$, $Q_{S_1S_2}(\theta)$, $Q_{S_2S_2}(\theta)$ evaluated at $\theta = \theta^\dagger + \lambda(\hat{\theta}_n - \theta^\dagger)$, $\lambda \in [0,1]$ in Appendix D.4, we prove that as $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, the following convergences hold uniformly in $\lambda \in [0,1]$:

(a) For $1 \le i \le N_\beta$, $N_{\beta_S} + 1 \le j \le N_\beta$,

$$\begin{aligned}&\big[\mathscr{I}_n\big(\theta^\dagger + \lambda(\hat{\theta}_n - \theta^\dagger)\big)\big]_{ij} \\ &\xrightarrow{\mathbb{P}_{\theta^\dagger}} \begin{cases} 2\int \big(\partial_i^\beta V_{R,0}(x, \beta_R^\dagger)\big)^\top a_R^{-1}(x, \sigma^\dagger)\, \partial_j^\beta V_{R,0}(x, \beta_R^\dagger)\, \nu_{\theta^\dagger}(dx), & N_{\beta_S} + 1 \le i, j \le N_\theta; \\ 0, & \text{(otherwise).}\end{cases}\end{aligned} \tag{54}$$

(b) For $1 \le i \le N_\beta$, $N_\beta + 1 \le j \le N_\theta$,

$$\big[\mathscr{I}_n\big(\theta^\dagger + \lambda(\hat{\theta}_n - \theta^\dagger)\big)\big]_{ij} \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0. \tag{55}$$

(c) For $N_\beta + 1 \le i, j \le N_\theta$,

$$\big[\mathscr{I}_n\big(\theta^\dagger + \lambda(\hat{\theta}_n - \theta^\dagger)\big)\big]_{ij} \xrightarrow{\mathbb{P}_{\theta^\dagger}} \int \mathrm{tr}\big(\partial_i^\sigma \Sigma(x, \theta^\dagger)\, \Lambda(x, \theta^\dagger) \partial_j^\sigma \Sigma(x, \theta^\dagger)\, \Lambda(x, \theta^\dagger)\big) \nu_{\theta^\dagger}(dx), \tag{56}$$

D.7.1. Proof of (54)

We consider the following three cases separately:

(a1) $1 \le i \le N_{\beta_{S_1}}$, $N_{\beta_S} + 1 \le j \le N_\beta$;

(a2) $N_{\beta_{S_1}} + 1 \le i \le N_{\beta_S}$, $N_{\beta_S} + 1 \le j \le N_\beta$;

(a3) $N_{\beta_S} + 1 \le i, j \le N_\beta$.

In case (a1), we have for $\theta = (\beta_{S_1}, \beta_{S_2}, \beta_R, \sigma) \in \Theta$,

$$\big[\mathscr{I}_n(\theta)\big]_{ij} = \frac{1}{n}\sum_{m=1}^{n} H_{ij}(X_{m-1},\theta) + \frac{1}{n}\sum_{m=1}^{n} R(\Delta_n, X_{m-1},\theta) + \frac{1}{n}\sum_{m=1}^{n}\sum_{1\le k\le N_{S_1}} R_k(1, X_{m-1},\theta)\big(V^k_{S_1,0}(X_{S,m-1},\beta^\dagger_{S_1}) - V^k_{S_1,0}(X_{S,m-1},\beta_{S_1})\big), \tag{57}$$

for some $R_k, R \in \mathcal{S}$, $1 \le k \le N_{S_1}$, where we have defined $H_{ij}(x,\theta)$, $(x,\theta) \in \mathbb{R}^N \times \Theta$ as:

$$H_{ij}(x,\theta) = \tfrac{1}{3}\big(\partial^\beta_i V_{S_1,0}(x_{S_1},\beta_{S_1})\big)^\top \Lambda_{S_1S_1}(x,\theta)\,\partial^\beta_j \mathcal{L}^2 V_{S_1,0}(x,\theta) + \big(\partial^\beta_i V_{S_1,0}(x_{S_1},\beta_{S_1})\big)^\top \Lambda_{S_1S_2}(x,\theta)\,\partial^\beta_j \mathcal{L} V_{S_2,0}(x,\theta) + 2\big(\partial^\beta_i V_{S_1,0}(x_{S_1},\beta_{S_1})\big)^\top \Lambda_{S_1R}(x,\theta)\,\partial^\beta_j V_{R,0}(x,\beta_R).$$

Noticing that for $N_{\beta_S}+1 \le j \le N_\beta$,

$$\partial^\beta_j \mathcal{L}V_{S_2,0}(x,\theta) = \partial^\top_{x_R} V_{S_2,0}(x,\beta_{S_2})\partial^\beta_j V_{R,0}(x,\beta_R), \quad \partial^\beta_j \mathcal{L}^2 V_{S_1,0}(x,\theta) = \partial^\top_{x_{S_2}} V_{S_1,0}(x_S,\beta_{S_1})\partial^\top_{x_R} V_{S_2,0}(x,\beta_{S_2})\partial^\beta_j V_{R,0}(x,\beta_R),$$

we have $H_{ij}(x,\theta) = 0$ for any $(x,\theta) \in \mathbb{R}^N \times \Theta$ since

$$H_{ij}(x,\theta) = 2\big(\partial^\beta_i V_{S_1,0}(x,\beta_{S_1})\big)^\top \widetilde{H}_{ij}(x,\theta)\,\partial^\beta_j V_{R,0}(x,\beta_R),$$

where

$$\widetilde{H}_{ij}(x,\theta) = \big\{\Lambda_{S_1S_1}(x,\theta)\Sigma_{S_1R}(x,\theta) + \Lambda_{S_1S_2}(x,\theta)\Sigma_{S_2R}(x,\theta) + \Lambda_{S_1R}(x,\theta)\Sigma_{RR}(x,\sigma)\big\}a_R^{-1}(x,\sigma) = \mathbf{0}_{N_{S_1}\times N_R}.$$

Thus, due to the consistency of the estimator and Lemma 2, we immediately obtain from (57) that if $n\to\infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$[\mathscr{I}_n\big(\theta^\dagger + \lambda(\hat\theta_n - \theta^\dagger)\big)]_{ij} \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

uniformly in $\lambda \in [0,1]$ for $1 \le i \le N_{\beta_{S_1}}$, $N_{\beta_S}+1 \le j \le N_\beta$.

Subsequently, we consider the case (a2). We have

$$\big[\mathscr{I}_n(\theta)\big]_{ij} = \frac{1}{n}\sum_{m=1}^{n} \widetilde{H}_{ij,1}(X_{m-1},\theta) + \frac{1}{n}\sum_{m=1}^{n} \widetilde{H}_{ij,2}(X_{m-1},\theta) + \frac{1}{n}\sum_{m=1}^{n}\sum_{1\le k\le N_{S_1}} R_k(1,X_{m-1},\theta)\eta^k_{S_1,m-1}(\Delta_n,\theta) + \frac{1}{n}\sum_{m=1}^{n}\sum_{N_{S_1}+1\le k\le N_S} R_k(1,X_{m-1},\theta)\big(V^k_{S,0}(X_{m-1},\beta^\dagger_S) - V^k_{S,0}(X_{m-1},\beta_S)\big) + \frac{1}{n}\sum_{m=1}^{n} R(\Delta_n, X_{m-1},\theta),$$

for some $R_k, R \in \mathcal{S}$, $1 \le k \le N_S$, where we have defined $\widetilde{H}_{ij,k}(x,\theta)$, for $(x,\theta) \in \mathbb{R}^N \times \Theta$, $k = 1,2$, as:

$$\begin{aligned}
\widetilde{H}_{ij,1}(x,\theta) &\equiv \tfrac{1}{6}\big(\partial^\beta_i \mathcal{L}V_{S_1,0}(x,\theta)\big)^\top \Lambda_{S_1S_1}(x,\theta)\,\partial^\beta_j \mathcal{L}^2 V_{S_1,0}(x,\theta) + \tfrac{1}{2}\big(\partial^\beta_i \mathcal{L}V_{S_1,0}(x,\theta)\big)^\top \Lambda_{S_1S_2}(x,\theta)\,\partial^\beta_j \mathcal{L}V_{S_2,0}(x,\theta) \\
&\quad + \big(\partial^\beta_i \mathcal{L}V_{S_1,0}(x,\theta)\big)^\top \Lambda_{S_1R}(x,\theta)\,\partial^\beta_j V_{R,0}(x,\beta_R); \\
\widetilde{H}_{ij,2}(x,\theta) &\equiv \big(\tfrac{1}{3}\partial^\beta_i V_{S_2,0}(x,\beta_{S_2})\big)^\top \Lambda_{S_2S_1}(x,\theta)\,\partial^\beta_j \mathcal{L}^2 V_{S_1,0}(x,\theta) + \big(\partial^\beta_i V_{S_2,0}(x,\beta_{S_2})\big)^\top \Lambda_{S_2S_2}(x,\theta)\,\partial^\beta_j \mathcal{L}V_{S_2,0}(x,\theta) \\
&\quad + 2\big(\partial^\beta_i V_{S_2,0}(x,\beta_{S_2})\big)^\top \Lambda_{S_2R}(x,\theta)\,\partial^\beta_j V_{R,0}(x,\beta_R).
\end{aligned}$$

We then have $\widetilde{H}_{ij,k}(x,\theta) = 0$ for any $(x,\theta) \in \mathbb{R}^N \times \Theta$, $k = 1,2$, from the same argument as in case (a1). Thus, making use of Lemma 2, convergence (43) with condition (C2) and the consistency of the estimator, we obtain (54) in case (a2).

Finally, we consider the case (a3). We have

$$\big[\mathscr{I}_n(\theta)\big]_{ij} = \frac{1}{n}\sum_{k=1}^{n} \bar{H}_{ij,1}(X_{k-1},\theta) + \frac{1}{n}\sum_{k=1}^{n} \bar{H}_{ij,2}(X_{k-1},\theta) + \frac{1}{n}\sum_{k=1}^{n} \bar{H}_{ij,3}(X_{k-1},\theta) + \frac{1}{n\sqrt{\Delta_n}}\sum_{k=1}^{n} R(1,X_{k-1},\theta)\,m_k(\Delta_n,\theta^\dagger) + \frac{1}{n\sqrt{\Delta_n}}\sum_{k=1}^{n}\big(m_k(\Delta_n,\theta) - m_k(\Delta_n,\theta^\dagger)\big)^\top \Lambda(X_{k-1},\theta)\,\partial^\beta_{(i,j)} v(X_{k-1},\theta) + \frac{1}{n}\sum_{k=1}^{n} \widetilde{R}(\sqrt{\Delta_n}, X_{k-1},\theta),$$

for R, $\widetilde{R} \in \mathcal{S}$, where we have set, for $(x,\theta) \in \mathbb{R}^N \times \Theta$,

$$\begin{aligned}
\bar{H}_{ij,1}(x,\theta) \;&\equiv \tfrac{1}{18}\big(\partial^\beta_i \mathcal{L}^2 V_{S_1,0}(x,\theta)\big)^\top \Lambda_{S_1S_1}(x,\theta)\,\partial^\beta_j \mathcal{L}^2 V_{S_1,0}(x,\theta) \\
&\quad + \tfrac{1}{6}\big(\partial^\beta_i \mathcal{L}^2 V_{S_1,0}(x,\theta)\big)^\top \Lambda_{S_1S_2}(x,\theta)\,\partial^\beta_j \mathcal{L}V_{S_2,0}(x,\theta) \\
&\quad + \tfrac{1}{3}\big(\partial^\beta_i \mathcal{L}^2 V_{S_1,0}(x,\theta)\big)^\top \Lambda_{S_1R}(x,\theta)\,\partial^\beta_j V_{R,0}(x,\beta_R);
\end{aligned} \tag{58}$$

$$\begin{aligned}
\bar{H}_{ij,2}(x,\theta) \;&\equiv \tfrac{1}{6}\big(\partial^\beta_i \mathcal{L}V_{S_2,0}(x,\theta)\big)^\top \Lambda_{S_2S_1}(x,\theta)\,\partial^\beta_j \mathcal{L}^2 V_{S_1,0}(x,\theta) \\
&\quad + \tfrac{1}{2}\big(\partial^\beta_i \mathcal{L}V_{S_2,0}(x,\theta)\big)^\top \Lambda_{S_2S_2}(x,\theta)\,\partial^\beta_j \mathcal{L}V_{S_2,0}(x,\theta) \\
&\quad + \big(\partial^\beta_i \mathcal{L}V_{S_2,0}(x,\theta)\big)^\top \Lambda_{S_2R}(x,\theta)\,\partial^\beta_j V_{R,0}(x,\beta_R);
\end{aligned} \tag{59}$$

$$\begin{aligned}\bar{H}_{ij,3}(x,\theta) \quad &\equiv \tfrac{1}{3}\big(\partial_i^{\beta} V_{R,0}(x,\beta_R)\big)^{\top} \Lambda_{RS_1}(x,\theta)\,\partial_j^{\beta}\mathcal{L}^2 V_{S_1,0}(x,\theta) \\ &+\big(\partial_i^{\beta} V_{R,0}(x,\beta_R)\big)^{\top} \Lambda_{RS_2}(x,\theta)\,\partial_j^{\beta}\mathcal{L} V_{S_2,0}(x,\theta) \\ &+2\big(\partial_i^{\beta} V_{R,0}(x,\beta_R)\big)^{\top} \Lambda_{RR}(x,\theta)\,\partial_j^{\beta} V_{R,0}(x,\beta_R),\end{aligned} \tag{60}$$

and

$$v(x,\theta) \equiv \Big[\tfrac{1}{6}\mathcal{L}^2 V_{S_1,0}(x,\theta)^{\top},\ \tfrac{1}{2}\mathcal{L} V_{S_2,0}(x,\theta)^{\top},\ V_{R,0}(x,\beta_R)^{\top}\Big]^{\top}.$$

Notice that for any $(x,\theta)\in\mathbb{R}^N\times\Theta$,

$$\bar{H}_{ij,1}(x,\theta)=0, \qquad \bar{H}_{ij,2}(x,\theta)=0, \qquad \bar{H}_{ij,3}(x,\theta)=2\big(\partial_i^{\beta} V_{R,0}(x,\beta_R)\big)^{\top} a_R^{-1}(x,\sigma)\,\partial_j^{\beta} V_{R,0}(x,\beta_R).$$

Furthermore, it follows that

$$\begin{aligned}&\frac{1}{n\sqrt{\Delta_n}}\sum_{k=1}^{n}\big(m_k(\Delta_n,\theta)-m_k(\Delta_n,\theta^{\dagger})\big)^{\top}\Lambda(X_{k-1},\theta)\partial^{\beta}_{(i,j)}v(X_{k-1},\theta) \\ &=\frac{1}{n}\sum_{k=1}^{n}\big(V_{R,0}(X_{k-1},\beta_R^{\dagger})-V_{R,0}(X_{k-1},\beta_R)\big)^{\top} a_R^{-1}(X_{k-1},\sigma)\partial^{\beta}_{(i,j)}V_{R,0}(X_{k-1},\beta_R),\end{aligned}$$

where we made use of similar arguments in the proof of Lemma 6 in Appendix A.2 for the term $\Lambda(X_{k-1},\theta)\partial^{\beta}_{(i,j)}v(X_{k-1},\theta)$. Hence, exploiting Lemmas 2, 3 and the consistency of estimator $\hat{\theta}_n$, we obtain that:

$$\Big[\mathscr{I}_n\big(\theta^{\dagger}+\lambda(\hat{\theta}_n-\theta^{\dagger})\big)\Big]_{ij}\xrightarrow{\mathbb{P}_{\theta^{\dagger}}} 2\int\big(\partial_i^{\beta} V_{R,0}(x,\beta_R^{\dagger})\big)^{\top}\, a_R^{-1}(x,\sigma^{\dagger})\,\partial_j^{\beta} V_{R,0}(x,\beta_R^{\dagger})\,\nu_{\theta^{\dagger}}(dx),$$

as $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$, for $N_\beta+1\le i,j\le N_\theta$. The proof of (54) is now complete.

D.7.2. Proof of (55)

We show (55) when $1\le i\le N_{\beta_{S_1}}$ and $N_\beta+1\le j\le N_\theta$. The convergence for the other cases can be deduced from a similar argument used in the proof of (54) so we omit the proof. We have

$$\begin{aligned}\big[\mathscr{I}_n(\theta)\big]_{ij} &= \frac{\sqrt{\Delta_n^3}}{n}\sum_{k=1}^{n}\sum_{1\le k_1,k_2\le N} R_{k_1k_2}(1,X_{k-1},\theta)m_k^{k_1}(\Delta_n,\theta^{\dagger})m_k^{k_2}(\Delta_n,\theta^{\dagger}) \\ &+\frac{\sqrt{\Delta_n^3}}{n}\sum_{k=1}^{n}\sum_{1\le k_1,k_2\le N}\Big\{\widetilde{R}_{k_1k_2}(1,X_{k-1},\theta)\big(m_k^{k_1}(\Delta_n,\theta)-m_k^{k_1}(\Delta_n,\theta^{\dagger})\big)\big(m_k^{k_2}(\Delta_n,\theta)-m_k^{k_2}(\Delta_n,\theta^{\dagger})\big)\Big\} \\ &+\frac{1}{n}\sum_{k=1}^{n}\sum_{1\le k_1\le N} R_{k_1}(1,X_{k-1},\theta)\big(m_k^{k_1}(\Delta_n,\theta)-m_k^{k_1}(\Delta_n,\theta^{\dagger})\big)+\frac{1}{n}\sum_{k=1}^{n}R(\sqrt{\Delta_n},X_{k-1},\theta),\end{aligned}$$

for some $R_{k_1k_2},\widetilde{R}_{k_1k_2},R_{k_1},R\in\mathcal{S}$. Thus, we immediately obtain (55) for $1\le i\le N_{\beta_{S_1}}$, $N_\beta+1\le j\le N_\theta$ from Lemma 2, 3 and (43)–(44).

D.7.3. Proof of (56)

It holds that for $N_\beta+1\le i,j\le N_\theta$, $\theta=(\beta_{S_1},\beta_{S_2},\beta_R,\sigma)\in\Theta$,

$$\begin{aligned}\big[\mathscr{I}_n(\theta)\big]_{ij} &= \frac{1}{n}\sum_{k=1}^{n} m_k(\Delta_n,\theta^{\dagger})^{\top}\,\partial^{\sigma}_{(i,j)}\Lambda(X_{k-1},\theta)\,m_k(\Delta_n,\theta^{\dagger})+\frac{1}{n}\sum_{k=1}^{n}\partial^{\sigma}_{(i,j)}\log\big|\Sigma(X_{k-1},\theta)\big| \\ &+\frac{1}{n}\sum_{k=1}^{n}\sum_{1\le k_1,k_2\le N} R_{k_1k_2}(1,X_{k-1},\theta)m_k^{k_1}(\Delta_n,\theta^{\dagger})\big(m_k^{k_2}(\Delta_n,\theta^{\dagger})-m_k^{k_2}(\Delta_n,\theta)\big) \\ &+\frac{1}{n}\sum_{k=1}^{n}\sum_{1\le k_1,k_2\le N}\Big\{\widetilde{R}_{k_1k_2}(1,X_{k-1},\theta)\big(m_k^{k_1}(\Delta_n,\theta^{\dagger})-m_k^{k_1}(\Delta_n,\theta)\big)\big(m_k^{k_2}(\Delta_n,\theta^{\dagger})-m_k^{k_2}(\Delta_n,\theta)\big)\Big\} \\ &+\frac{1}{n}\sum_{k=1}^{n}\sum_{1\le k_1\le N} R_{k_1}(\sqrt{\Delta_n},X_{k-1},\theta)\big(m_k^{k_1}(\Delta_n,\theta^{\dagger})-m_k^{k_1}(\Delta_n,\theta)\big)+\frac{1}{n}\sum_{k=1}^{n}R(\Delta_n,X_{k-1},\theta),\end{aligned}$$

for some $R_{k_1k_2},\widetilde{R}_{k_1k_2},R_{k_1},R\in\mathcal{S}$. Making use of Lemmas 2–3, (43), (44) and the consistency of the estimator, we obtain as $n\to\infty$, $\Delta_n\to 0$ and $n\Delta_n\to\infty$,

$$\begin{aligned}\Big[\mathscr{I}_n\big(\theta^{\dagger}+\lambda(\hat{\theta}_n-\theta^{\dagger})\big)\Big]_{ij} &\xrightarrow{\mathbb{P}_{\theta^{\dagger}}}\int\Big\{\mathrm{tr}\big(\partial^{\sigma}_{(i,j)}\Lambda(x,\theta^{\dagger})\,\Sigma(x,\theta^{\dagger})\big)+\partial^{\sigma}_{(i,j)}\log\big|\Sigma(x,\theta^{\dagger})\big|\Big\}\nu_{\theta^{\dagger}}(dx) \\ &=\int\mathrm{tr}\big(\partial_i^{\sigma}\Sigma(x,\theta^{\dagger})\,\Lambda(x,\theta^{\dagger})\partial_j^{\sigma}\Sigma(x,\theta^{\dagger})\,\Lambda(x,\theta^{\dagger})\big)\nu_{\theta^{\dagger}}(dx),\end{aligned}$$

uniformly in $\lambda \in [0,1]$, where we applied the following two formulae to the above equation:

$$\partial^{\sigma}_{(i,j)} \log \left|\Sigma(x,\theta^{\dagger})\right| = -\mathrm{tr}\left(\partial^{\sigma}_{(i,j)} \Lambda(x,\theta^{\dagger})\, \Sigma(x,\theta^{\dagger})\right) - \mathrm{tr}\left(\partial^{\sigma}_{i} \Sigma(x,\theta^{\dagger}) \partial^{\sigma}_{j} \Lambda(x,\theta^{\dagger})\right);$$
$$\mathrm{tr}\left(\partial^{\sigma}_{i} \Sigma(x,\theta^{\dagger}) \partial^{\sigma}_{j} \Lambda(x,\theta^{\dagger})\right) = -\mathrm{tr}\left(\partial^{\sigma}_{i} \Sigma(x,\theta^{\dagger})\, \Lambda(x,\theta^{\dagger}) \partial^{\sigma}_{j} \Sigma(x,\theta^{\dagger})\, \Lambda(x,\theta^{\dagger})\right).$$

The proof is now complete.

D.8. Proof of Lemma 14

We write $\mathscr{C}^k_n(\theta) = \sum_{i=1}^n \zeta^k_i(\theta),\ \theta \in \Theta,\ 1 \le k \le N_\theta$, where we have set:

$$\zeta^k_i(\theta) \equiv \left[M_n^{-1}\right]_{kk} \times \partial^{\theta}_k \left\{ m_i(\Delta_n,\theta)^{\top} \Lambda(X_{i-1},\theta) m_i(\Delta_n,\theta) + \log|\Sigma(X_{i-1},\theta)| \right\}.$$

To prove the assertion, it suffices to show, from Theorem 3.2 and 3.4 in [17], that:

(i) If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$ with $\Delta_n = o(n^{-1/2})$, then

$$\sum_{i=1}^n \mathbb{E}_{\theta^{\dagger}} \left[\zeta^k_i(\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] \xrightarrow{\mathbb{P}_{\theta^{\dagger}}} 0, \quad 1 \le k \le N_\theta. \tag{61}$$

(ii) If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$\sum_{i=1}^n \mathbb{E}_{\theta^{\dagger}} \left[\zeta^{k_1}_i(\theta^{\dagger}) \zeta^{k_2}_i(\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] \xrightarrow{\mathbb{P}_{\theta^{\dagger}}} 4[\Gamma(\theta^{\dagger})]_{k_1 k_2}, \quad 1 \le k_1, k_2 \le N_\theta. \tag{62}$$

(iii) If $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$\sum_{i=1}^n \mathbb{E}_{\theta^{\dagger}} \left[\zeta^{k_1}_i(\theta^{\dagger}) \zeta^{k_2}_i(\theta^{\dagger}) \zeta^{k_3}_i(\theta^{\dagger}) \zeta^{k_4}_i(\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] \xrightarrow{\mathbb{P}_{\theta^{\dagger}}} 0, \quad 1 \le k_1, k_2, k_3, k_4 \le N_\theta. \tag{63}$$

In what follows, we will check convergences (61) and (62). One can prove (63) following similar arguments and by noticing that the left-hand-side of (63) involves $1/n^2$.

D.8.1. Proof of (61)

We recall (32) and (33) that are immediately obtained from the definition of $m_i(\Delta_n,\theta^{\dagger})$ in (28), that is, for $1 \le k_1, k_2 \le N$,

$$\mathbb{E}_{\theta^{\dagger}} \left[m^{k_1}_i(\Delta_n,\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] = R_1(\sqrt{\Delta_n^3}, X_{i-1}, \theta^{\dagger}); \tag{64}$$
$$\mathbb{E}_{\theta^{\dagger}} \left[m^{k_1}_i(\Delta_n,\theta^{\dagger}) m^{k_2}_i(\Delta_n,\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] = [\Sigma(X_{i-1},\theta^{\dagger})]_{k_1 k_2} + R_2(\Delta_n, X_{i-1}, \theta^{\dagger}) \tag{65}$$

for $R_1, R_2 \in \mathcal{S}$. We then write $\zeta^k_i(\theta),\ \theta \in \Theta,\ 1 \le i \le n,\ 1 \le k \le N_\theta$ as $\zeta^k_i(\theta) = \zeta^k_{i,1}(\theta) + \zeta^k_{i,2}(\theta)$, where we have set:

$$\zeta^k_{i,1}(\theta) = 2\left[M_n^{-1}\right]_{kk} \times \left\{ \left(\partial^{\theta}_k m_i(\Delta_n,\theta)\right)^{\top} \Lambda(X_{i-1},\theta)\, m_i(\Delta_n,\theta) \right\};$$
$$\zeta^k_{i,2}(\theta) = \left[M_n^{-1}\right]_{kk} \times \left\{ \partial^{\theta}_k \log|\Sigma(X_{i-1},\theta)| + m_i(\Delta_n,\theta)^{\top} \partial^{\theta}_k \Lambda(X_{i-1},\theta)\, m_i(\Delta_n,\theta) \right\}.$$

Exploiting (64), we obtain

$$\mathbb{E}_{\theta^{\dagger}} \left[\zeta^k_{i,1}(\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] = \tfrac{1}{\sqrt{n}} \sum_{j=1}^N R^j_k(1, X_{i-1}, \theta^{\dagger})\, \mathbb{E}_{\theta^{\dagger}} \left[m^j_i(\Delta_n,\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] = \tfrac{1}{n} \widetilde{R}^j_k(\sqrt{n\Delta_n^3}, X_{i-1}, \theta^{\dagger}),$$

for $R^j_k,\ \widetilde{R}^j_k \in \mathcal{S}$. Thus, from Lemma 2, we obtain

$$\sum_{i=1}^n \mathbb{E}_{\theta^{\dagger}} \left[\zeta^k_{i,1}(\theta^{\dagger}) | \mathcal{F}_{t_{i-1}} \right] = \tfrac{1}{n} \sum_{i=1}^n \widetilde{R}^j_k(\sqrt{n\Delta_n^3}, X_{i-1}, \theta^{\dagger}) \xrightarrow{\mathbb{P}_{\theta^{\dagger}}} 0,$$

if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$ and $\Delta_n = o(n^{-1/2})$. Next, we consider the second term $\zeta^k_{i,2}(\theta)$. First, notice that for $N_{\beta_S} + 1 \le k \le N_\beta$, $\zeta^k_{i,2}(\theta^{\dagger}) = 0$, since $\Sigma(X_{i-1},\theta)$ and $\Lambda(X_{i-1},\theta)$ are independent of $\beta_R \in \Theta_{\beta_R}$. For $1 \le k \le N_{\beta_S}$ and $N_\beta + 1 \le k \le N_\theta$, we apply (65) to obtain

$$\begin{aligned}\mathbb{E}_{\theta^{\dagger}}[\zeta^k_{i,2}(\theta^{\dagger}) | \mathcal{F}_{t_{i-1}}] &= [M_n^{-1}]_{kk} \times \left\{ \partial^{\theta}_k \log|\Sigma(X_{i-1},\theta^{\dagger})| + \mathrm{tr}\left(\partial^{\theta}_k \Lambda(X_{i-1},\theta^{\dagger}) \Sigma(X_{i-1},\theta^{\dagger})\right) \right\} + \tfrac{1}{n} R_k(\sqrt{n\Delta_n^2}, X_{i-1}, \theta^{\dagger}) \\ &= \tfrac{1}{n} R_k(\sqrt{n\Delta_n^2}, X_{i-1}, \theta^{\dagger})\end{aligned}$$

for $R_k \in \mathcal{S}$, where we used:

$$\partial^{\theta}_k \log|\Sigma(x,\theta)| = -\mathrm{tr}\left(\partial^{\theta}_k \Lambda(x,\theta) \Sigma(x,\theta)\right), \quad (x,\theta) \in \mathbb{R}^N \times \Theta. \tag{66}$$

Thus, we have from Lemma 2 that if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$ with $\Delta_n = o(n^{-1/2})$, then

$$\sum_{i=1}^{n} \mathbb{E}_{\theta^\dagger}\left[\zeta_{i,2}^{k}(\theta^\dagger)|\mathcal{F}_{t_{i-1}}\right] = \frac{1}{n}\sum_{i=1}^{n} R_k(\sqrt{n\Delta_n^2}, X_{i-1}, \theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

and the proof of (61) is now complete.

D.8.2. Proof of (62)

For simplicity, we write

$$\mathscr{Y}_{k_1k_2}(\theta^\dagger) \equiv \sum_{i=1}^{n} \mathbb{E}_{\theta^\dagger}\left[\zeta_i^{k_1}(\theta^\dagger)\zeta_i^{k_2}(\theta^\dagger)|\mathcal{F}_{t_{i-1}}\right].$$

We have that for $1 \le k_1 \le N_{\beta_S}$, $N_{\beta_S}+1 \le k_2 \le N_\beta$, $N_\beta + 1 \le k_3 \le N_\theta$,

$$\begin{aligned}\sqrt{n}\zeta_i^{k_1}(\theta^\dagger) &= -2\,\mu_{k_1}(X_{i-1},\theta^\dagger)^\top \Lambda(X_{i-1},\theta^\dagger)\, m_i(\Delta_n,\theta^\dagger) \\ &\quad + \textstyle\sum_{1\le j_1,j_2\le N} R_{k_1}^{j_1j_2}(\sqrt{\Delta_n}, X_{i-1},\theta^\dagger) m_i^{j_1}(\Delta_n,\theta^\dagger) m_i^{j_2}(\Delta_n,\theta^\dagger) + \sum_{1\le j_1\le N} R_{k_1}^{j_1}(\sqrt{\Delta_n}, X_{i-1},\theta^\dagger) m_i^{j_1}(\Delta_n,\theta^\dagger);\end{aligned} \tag{67}$$

$$\sqrt{n}\zeta_i^{k_2}(\theta^\dagger) = -2\,\mu_{k_2}(X_{i-1},\theta^\dagger)^\top \Lambda(X_{i-1},\theta^\dagger) m_i(\Delta_n,\theta^\dagger); \tag{68}$$

$$\begin{aligned}\sqrt{n}\zeta_i^{k_3}(\theta^\dagger) &= m_i(\Delta_n,\theta^\dagger)^\top \left(\partial_{k_3}^{\theta}\Lambda(X_{i-1},\theta^\dagger)\right) m_i(\Delta_n,\theta^\dagger) + \partial_{k_3}^{\theta} \log|\Sigma(X_{i-1}\theta^\dagger)| \\ &\quad + \textstyle\sum_{1\le j_1,j_2\le N} R_{k_3}^{j_1j_2}(\Delta_n, X_{i-1},\theta^\dagger)\, m_i^{j_1}(\Delta_n,\theta^\dagger) m_i^{j_2}(\Delta_n,\theta^\dagger) + \sum_{1\le j_1,j_2\le N} R_{k_3}^{j_1}(\sqrt{\Delta_n}, X_{i-1},\theta^\dagger)\, m_i^{j_1}(\Delta_n,\theta^\dagger),\end{aligned} \tag{69}$$

for $R_{k_1}^{j_1j_2}$, $R_{k_1}^{j_1}$, $R_{k_3}^{j_1j_2}$, $R_{k_3}^{j_1} \in \mathcal{S}$, where we defined $\mu_k : \mathbb{R}^N \times \Theta \to \mathbb{R}^N$, $1 \le k \le N_\beta$ as:

$$\mu_k(x,\theta) = \begin{cases} \left[\left(\partial_k^\beta V_{S_1,0}(x_S,\beta_{S_1})\right)^\top, \mathbf{0}_{N_{S_2}}^\top, \mathbf{0}_{N_R}^\top\right]^\top, & 1 \le k \le N_{\beta_{S_1}}; \\ \left[\frac{1}{2}\left(\partial_k^\beta \mathcal{L}V_{S_1,0}(x,\theta)\right)^\top, \left(\partial_k^\beta V_{S_2,0}(x,\beta_{S_2})\right)^\top, \mathbf{0}_{N_R}^\top\right]^\top, & N_{\beta_{S_1}}+1 \le k \le N_{\beta_S}; \\ \left[\frac{1}{6}\left(\partial_k^\beta \mathcal{L}^2 V_{S_1,0}(x,\theta)\right)^\top, \frac{1}{2}\left(\partial_k^\beta \mathcal{L}V_{S_2,0}(x,\theta)\right)^\top, \left(\partial_k^\beta V_{R,0}(x,\beta_R)\right)^\top\right]^\top, & N_{\beta_S}+1 \le k \le N_\beta, \end{cases}$$

for $(x,\theta) \in \mathbb{R}^N \times \Theta$. From Lemma 2, 3 and (67), we have that for $1 \le k_1, k_2 \le N_\beta$,

$$\begin{aligned}\mathscr{Y}_{k_1k_2}(\theta^\dagger) &\xrightarrow{\mathbb{P}_{\theta^\dagger}} 4 \sum_{1\le j_1,j_2,j_3,j_4\le N} \int \mu_{k_1}^{j_1}(x,\theta^\dagger)\left[\Lambda(x,\theta^\dagger)\right]_{j_1j_2} \mu_{k_2}^{j_3}(x,\theta^\dagger)\left[\Lambda(x,\theta^\dagger)\right]_{j_3j_4}\left[\Sigma(x,\theta^\dagger)\right]_{j_2j_4} \nu_{\theta^\dagger}(dx) \\ &= 4\sum_{1\le j_1,j_2\le N}\int \mu_{k_1}^{j_1}(x,\theta^\dagger)\left[\Lambda(x,\theta^\dagger)\right]_{j_1j_2}\mu_{k_2}^{j_2}(x,\theta^\dagger)\nu_{\theta^\dagger}(dx) \equiv 4\int \widetilde{\mathscr{Y}}_{k_1k_2}(x,\theta^\dagger)\,\nu_{\theta^\dagger}(dx).\end{aligned}$$

We then have, for $1 \le k_1, k_2 \le N_{\beta_{S_1}}$,

$$\begin{aligned}\widetilde{\mathscr{Y}}_{k_1k_2}(x,\theta^\dagger) &= \left(\partial_{k_1}^\beta V_{S_1,0}(x_S,\beta_{S_1}^\dagger)\right)^\top \Lambda_{S_1S_1}(x,\theta^\dagger)\,\partial_{k_2}^\beta V_{S_1,0}(x_S,\beta_{S_1}^\dagger) \\ &= 720\left(\partial_{k_1}^\beta V_{S_1,0}(x_S,\beta_{S_1}^\dagger)\right)^\top a_{S_1}^{-1}(x,\theta^\dagger)\,\partial_{k_2}^\beta V_{S_1,0}(x_S,\beta_{S_1}^\dagger),\end{aligned} \tag{70}$$

for $N_{\beta_{S_1}}+1 \le k_1, k_2 \le N_{\beta_S}$,

$$\begin{aligned}\widetilde{\mathscr{Y}}_{k_1k_2}(x,\theta^\dagger) &= \tfrac{1}{4}\left(\partial_{k_1}^\beta \mathcal{L}V_{S_1,0}(x,\theta^\dagger)\right)^\top \Lambda_{S_1S_1}(x,\theta^\dagger)\,\partial_{k_2}^\beta \mathcal{L}V_{S_1,0}(x,\theta^\dagger) \\ &\quad + \tfrac{1}{2}\left(\partial_{k_1}^\beta \mathcal{L}V_{S_1,0}(x,\theta^\dagger)\right)^\top \Lambda_{S_1S_2}(x,\theta^\dagger)\,\partial_{k_2}^\beta V_{S_2,0}(x,\beta_{S_2}^\dagger) \\ &\quad + \tfrac{1}{2}\left(\partial_{k_1}^\beta V_{S_2,0}(x,\beta_{S_2}^\dagger)\right)^\top \Lambda_{S_2S_1}(x,\theta^\dagger)\,\partial_{k_2}^\beta \mathcal{L}V_{S_1,0}(x,\theta^\dagger) \\ &\quad + \left(\partial_{k_1}^\beta V_{S_2,0}(x,\beta_{S_2}^\dagger)\right)^\top \Lambda_{S_2S_2}(x,\theta^\dagger)\,\partial_{k_2}^\beta V_{S_2,0}(x,\beta_{S_2}^\dagger) \\ &= 12\left(\partial_{k_1}^\beta V_{S_2,0}(x,\beta_{S_2}^\dagger)\right)^\top a_{S_2}^{-1}(x,\theta^\dagger)\,\partial_{k_2}^\beta V_{S_2,0}(x,\beta_{S_2}^\dagger),\end{aligned} \tag{71}$$

and for $N_{\beta_S}+1 \le k_1, k_2 \le N_\beta$,

$$\widetilde{\mathscr{Y}}_{k_1k_2}(x,\theta^\dagger) = \frac{1}{2}\sum_{j=1}^{3}\bar{H}_{k_1k_2,j}(x,\theta^\dagger) = \left(\partial_{k_1}^\beta V_{R,0}(x,\beta_R^\dagger)\right)^\top a_R^{-1}(x,\sigma^\dagger)\,\partial_{k_2}^\beta V_{R,0}(x,\beta_R^\dagger),$$

where $\bar{H}_{k_1k_2,j}$ $1 \le j \le 3$ are defined as (58), (59) and (60). Note that we made use of Lemma 4, 5 to derive (70)–(71). For $1 \le k_1 \le N_\beta$, $N_\beta + 1 \le k_2 \le N_\theta$, it follows from (67), (68) and (69) that

$$\mathscr{Y}_{k_1k_2}(\theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} 0,$$

if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$. For $N_{\beta_R}+1 \le k_1, k_2 \le N_\theta$, it follows that if $n \to \infty$, $\Delta_n \to 0$ and $n\Delta_n \to \infty$, then

$$\mathscr{Y}_{k_1k_2}(\theta^\dagger) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \widetilde{\mathscr{Y}}_{k_1k_2,1}(\theta^\dagger) + \widetilde{\mathscr{Y}}_{k_1k_2,2}(\theta^\dagger) + \widetilde{\mathscr{Y}}_{k_1k_2,3}(\theta^\dagger),$$

where we have set:

$$\widetilde{\mathscr{Y}}_{k_1k_2,1}(\theta^\dagger) \equiv \sum_{1\le j_1,j_2,j_3,j_4\le N} \int [\partial^\theta_{k_1} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\partial^\theta_{k_2} \Lambda(x,\theta^\dagger)]_{j_3j_4}\mathscr{W}_{j_1j_2j_3j_4}(x,\theta^\dagger)\, \nu_{\theta^\dagger}(dx);$$

$$\begin{aligned}\widetilde{\mathscr{Y}}_{k_1k_2,2}(\theta^\dagger) &\equiv \sum_{1\le j_1,j_2\le N} \int \Big\{ [\partial^\theta_{k_1} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\Sigma(x,\theta^\dagger)]_{j_1j_2}\partial^\theta_{k_2} \log|\Sigma(x,\theta^\dagger)| \\ &\qquad + [\partial^\theta_{k_2} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\Sigma(x,\theta^\dagger)]_{j_1j_2}\partial^\theta_{k_1} \log|\Sigma(x,\theta^\dagger)| \Big\}\nu_{\theta^\dagger}(dx) \\ &= -\sum_{1\le j_1,j_2,j_3,j_4\le N} \int \Big\{ [\partial^\theta_{k_1} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\Sigma(x,\theta^\dagger)]_{j_1j_2}[\partial^\theta_{k_2} \Lambda(x,\theta^\dagger)]_{j_3j_4}[\Sigma(x,\theta^\dagger)]_{j_3j_4} \\ &\qquad + [\partial^\theta_{k_2} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\Sigma(x,\theta^\dagger)]_{j_1j_2}[\partial^\theta_{k_1} \Lambda(x,\theta^\dagger)]_{j_3j_4}[\Sigma(x,\theta^\dagger)]_{j_3j_4} \Big\}\nu_{\theta^\dagger}(dx);\end{aligned}$$

$$\begin{aligned}\widetilde{\mathscr{Y}}_{k_1k_2,3}(\theta^\dagger) &\equiv \int \partial^\theta_{k_1} \log|\Sigma(x,\theta^\dagger)|\partial^\theta_{k_2} \log|\Sigma(x,\theta^\dagger)|\nu_{\theta^\dagger}(dx) \\ &= \sum_{1\le j_1,j_2,j_3,j_4\le N} \int [\partial^\theta_{k_1} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\Sigma(x,\theta^\dagger)]_{j_1j_2}\,[\partial^\theta_{k_2} \Lambda(x,\theta^\dagger)]_{j_3j_4}\,[\Sigma(x,\theta^\dagger)]_{j_3j_4}\nu_{\theta^\dagger}(dx),\end{aligned}$$

with $\mathscr{W}_{j_1j_2j_3j_4} : \mathbb{R}^N \times \Theta \to \mathbb{R}$ defined as follows, for $(x,\theta) \in \mathbb{R}^N \times \Theta$,

$$\mathscr{W}_{j_1j_2j_3j_4}(x,\theta) = [\Sigma(x,\theta)]_{j_1j_2}[\Sigma(x,\theta)]_{j_3j_4} + [\Sigma(x,\theta)]_{j_1j_3}[\Sigma(x,\theta)]_{j_2j_4} + [\Sigma(x,\theta)]_{j_1j_4}[\Sigma(x,\theta)]_{j_2j_3}.$$

Notice that we have used (66) in the computation of $\widetilde{\mathscr{Y}}_{k_1k_2,2}(\theta^\dagger)$ and $\widetilde{\mathscr{Y}}_{k_1k_2,3}(\theta^\dagger)$. Thus, we have

$$\begin{aligned}\sum_{m=1}^{3} \widetilde{\mathscr{Y}}_{k_1k_2,m}(\theta^\dagger) &= \sum_{1\le j_1,j_2,j_3,j_4\le N} \int \Big\{ [\partial^\theta_{k_1} \Lambda(x,\theta^\dagger)]_{j_1j_2}[\partial^\theta_{k_2} \Lambda(x,\theta^\dagger)]_{j_3j_4} \\ &\qquad \times \big([\Sigma(x,\theta^\dagger)]_{j_1j_3}[\Sigma(x,\theta^\dagger)]_{j_2j_4} + [\Sigma(x,\theta^\dagger)]_{j_1j_4}[\Sigma(x,\theta^\dagger)]_{j_2j_3}\big) \Big\}\nu_{\theta^\dagger}(dx) \\ &= 2\int \mathrm{tr}\big(\partial^\theta_{k_1}\Sigma(x,\theta^\dagger)\,\Lambda(x,\theta^\dagger)\,\partial^\theta_{k_2}\Sigma(x,\theta^\dagger)\,\Lambda(x,\theta^\dagger)\big)\,\nu_{\theta^\dagger}(dx),\end{aligned}$$

where in the second equality, we have used the following formula:

$$[\partial^\theta_k \Lambda(\theta^\dagger)]_{j_1j_2} = -\sum_{1\le j_3,j_4\le N} [\Lambda(x,\theta^\dagger)]_{j_1j_3}[\partial^\theta_k \Sigma(x,\theta^\dagger)]_{j_3j_4}[\Lambda(x,\theta^\dagger)]_{j_4j_2}, \quad N_\beta + 1 \le k \le N_\theta.$$

Furthermore, for other cases of $1 \le k_1, k_2 \le N_\theta$, it holds that:

$$\widetilde{\mathscr{Y}}_{k_1k_2}(x,\theta^\dagger) = 0,$$

where we have used Lemma 4 and that for $N_{\beta_S} + 1 \le k \le N_\beta$,

$$\Lambda(x,\theta)\,\mu_k(x,\theta) = \Lambda(x,\theta)\begin{bmatrix}\Sigma_{S_1R}(x,\theta)\\ \Sigma_{S_2R}(x,\theta)\\ \Sigma_{RR}(x,\theta)\end{bmatrix} a_R^{-1}(x,\sigma)\partial^\beta_k V_{R,0}(x,\beta_R) = \begin{bmatrix}\mathbf{0}_{N_{S_1}}\\ \mathbf{0}_{N_{S_2}}\\ a_R^{-1}(x,\sigma)\partial^\beta_k V_{R,0}(x,\beta_R)\end{bmatrix}.$$

for $x \in \mathbb{R}^N$, $\theta = (\beta_S, \beta_R, \sigma) \in \Theta$. The proof is now complete.

Appendix E. Proof for case study in Section 4.2.1

From (20), we have $(\hat{\sigma}_n)^2 = F_{1,n} + F_{2,n} + F_{3,n}$, where

$$F_{1,n} \equiv \frac{6}{\Delta_n^3} \times \frac{1}{n}\sum_{i=0}^{n-1}\big(\hat{p}_{i+1} - \hat{p}_i - s_i\Delta_n + s_i\tfrac{\Delta_n^2}{2}\big)^2;$$

$$F_{2,n} \equiv -\frac{6}{\Delta_n^2} \times \frac{1}{n}\sum_{i=0}^{n-1}\big(\hat{p}_{i+1} - \hat{p}_i - s_i\Delta_n + s_i\tfrac{\Delta_n^2}{2}\big)\big(s_{i+1} - s_i - s_i\Delta_n\big);$$

$$F_{3,n} \equiv \frac{2}{\Delta_n} \times \frac{1}{n}\sum_{i=0}^{n-1}\big(s_{i+1} - s_i - s_i\Delta_n\big)^2$$

with the hidden components $\hat{p}_i$ estimated by numerical differentiation:

$$\hat{p}_i = \frac{q_{i+1} - q_i}{\Delta_n}, \quad 0 \le i \le n.$$

Since the rough component s_t follows the linear SDE, the solution is explicitly given as: for $u \in [t_i, t_{i+1})$

$$s_u = s_{t_i}e^{-(u-t_i)} + \sigma^\dagger \int_{t_i}^{u} e^{-(u-v)}dB_v$$

under the true parameter $\sigma^\dagger$. Thus, we have

$$\hat{p}_i = p_i + \frac{s_i}{\Delta_n}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} e^{-(v-t_i)}dvdu + \frac{\sigma^\dagger}{\Delta_n}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} e^{-(v-w)}dB_w dvdu,$$

and then

$$\begin{aligned}\hat{p}_{i+1} - \hat{p}_i &= p_{i+1} - p_i + \frac{1}{\Delta_n}\left(\Delta_n - (1-e^{-\Delta_n})\right)(s_{i+1}-s_i) \\ &\quad + \frac{\sigma^\dagger}{\Delta_n}\int_{t_{i+1}}^{t_{i+2}}\int_{t_{i+1}}^{u}\int_{t_{i+1}}^{v} e^{-(v-w)}dB_w dvdu - \frac{\sigma^\dagger}{\Delta_n}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} e^{-(v-w)}dB_w dvdu,\end{aligned}$$

where we used:

$$\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} e^{-(v-t_i)}dvdu = \Delta_n - (1-e^{-\Delta_n}).$$

Since $\Delta_n \in [0,1)$ is assumed to be small, we use the Taylor expansion for the terms $e^{-\Delta_n}$, $e^{-(v-w)}$ and the stochastic Taylor expansion of (p_{i+1}, s_{i+1}) around (p_i, s_i) under the true parameter $\sigma^\dagger$ to obtain

$$\begin{aligned}\hat{p}_{i+1} - \hat{p}_i &= s_i\Delta_n + \sigma^\dagger B_{t_{i+1}-t_i}\frac{\Delta_n}{2} + \sigma^\dagger\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} dB_v du \\ &\quad + \frac{\sigma^\dagger}{\Delta_n}\int_{t_{i+1}}^{t_{i+2}}\int_{t_{i+1}}^{u}\int_{t_{i+1}}^{v} dB_w dvdu - \frac{\sigma^\dagger}{\Delta_n}\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dvdu + \Delta_n^2\xi_i,\end{aligned} \tag{72}$$

where the random variable $\{\xi_i\}_i$ appearing in the remainder term satisfies $\mathbb{E}|\xi_i|^2 \le C$ for some constant $C > 0$. We express the Gaussian random variables as:

$$\begin{aligned}&B_{t_{i+1}-t_i} = \sqrt{\Delta_n}\times z_i^{(1)}, \quad \int_{t_i}^{t_{i+1}}\int_{t_i}^{u} dB_v du = \sqrt{\Delta_n^3}\times\left(\frac{z_i^{(1)}}{2} + \frac{z_i^{(2)}}{2\sqrt{3}}\right), \\ &\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dvdu = \sqrt{\Delta_n^5}\times\left(\frac{z_i^{(1)}}{6} + \frac{z_i^{(2)}}{4\sqrt{3}} + \frac{z_i^{(3)}}{12\sqrt{5}}\right),\end{aligned}$$

where $\{z_i^{(j)}\}_{0\le i\le n+1,\, j=1,2,3}$ is an i.i.d. sequence of standard normal random variables so that it holds

$$\begin{aligned}&\mathbb{E}\left[(B_{t_{i+1}-t_i})^2\right] = \Delta_n, \quad \mathbb{E}\left[B_{t_{i+1}-t_i}\times\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} dB_v du\right)\right] = \frac{\Delta_n^2}{2}; \\ &\mathbb{E}\left[B_{t_{i+1}-t_i}\times\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dvdu\right)\right] = \frac{\Delta_n^3}{6}, \quad \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} dB_v du\right)^2\right] = \frac{\Delta_n^3}{3}; \\ &\mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u} dB_v du\right)\times\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dvdu\right)\right] = \frac{\Delta_n^4}{8}, \quad \mathbb{E}\left[\left(\int_{t_i}^{t_{i+1}}\int_{t_i}^{u}\int_{t_i}^{v} dB_w dvdu\right)^2\right] = \frac{\Delta_n^5}{20}.\end{aligned}$$

Then, (72) is written as:

$$\begin{aligned}\hat{p}_{i+1} - \hat{p}_i &= s_i\Delta_n + \sigma^\dagger\sqrt{\Delta_n^3}\frac{z_i^{(1)}}{2} + \sigma^\dagger\sqrt{\Delta_n^3}\left(\frac{z_i^{(1)}}{2} + \frac{z_i^{(2)}}{2\sqrt{3}}\right) \\ &\qquad + \sigma^\dagger\sqrt{\Delta_n^3}\left(\frac{z_{i+1}^{(1)}}{6} + \frac{z_{i+1}^{(2)}}{4\sqrt{3}} + \frac{z_{i+1}^{(3)}}{12\sqrt{5}}\right) - \sigma^\dagger\sqrt{\Delta_n^3}\left(\frac{z_i^{(1)}}{6} + \frac{z_i^{(2)}}{4\sqrt{3}} + \frac{z_i^{(3)}}{12\sqrt{5}}\right) + \Delta_n^2\xi_i \\ &= s_i\Delta_n + \sigma^\dagger\sqrt{\Delta_n^3}\left\{\left(\frac{z_{i+1}^{(1)}}{6} + \frac{z_{i+1}^{(2)}}{4\sqrt{3}} + \frac{z_{i+1}^{(3)}}{12\sqrt{5}}\right) + \left(\frac{5z_i^{(1)}}{6} + \frac{z_i^{(2)}}{4\sqrt{3}} - \frac{z_i^{(3)}}{12\sqrt{5}}\right)\right\} + \Delta_n^2\xi_i.\end{aligned} \tag{73}$$

From the ergodicity of the process $\{s_t\}_t$ and (73), we have that, as $n\to\infty$, $\Delta_n \to 0$ and $n\Delta_n\to\infty$,

$$F_{1,n} = \frac{6(\sigma^\dagger)^2}{n}\sum_{i=0}^{n-1}\left(\frac{z_{i+1}^{(1)}}{6} + \frac{z_{i+1}^{(2)}}{4\sqrt{3}} + \frac{z_{i+1}^{(3)}}{12\sqrt{5}} + \frac{5}{6}z_i^{(1)} + \frac{z_i^{(2)}}{4\sqrt{3}} - \frac{z_i^{(3)}}{12\sqrt{5}}\right)^2 + \frac{1}{n}\sum_{i=0}^{n} R_i^{(1)}(\Delta_n) \xrightarrow{\mathbb{P}_{\theta^\dagger}} \frac{23}{5}(\sigma^\dagger)^2; \tag{74}$$

$$F_{2,n} = -\frac{6(\sigma^\dagger)^2}{n}\sum_{i=0}^{n-1}\left(\frac{z_{i+1}^{(1)}}{6} + \frac{z_{i+1}^{(2)}}{4\sqrt{3}} + \frac{z_{i+1}^{(3)}}{12\sqrt{5}} + \frac{5}{6}z_i^{(1)} + \frac{z_i^{(2)}}{4\sqrt{3}} - \frac{z_i^{(3)}}{12\sqrt{5}}\right)z_i^{(1)} + \frac{1}{n}\sum_{i=0}^{n} R_i^{(2)}(\Delta_n) \xrightarrow{\mathbb{P}_{\theta^\dagger}} -5(\sigma^\dagger)^2, \tag{75}$$

where each $\{R_i^{(1)}(\Delta_n)\}_i$ and $\{R_i^{(2)}(\Delta_n)\}_i$ is sequence of random variables such that for $0\le i\le n$, $j=1,2$,

$$\{\mathbb{E}[|R_i^{(j)}(\Delta_n)|^2]\} \le C\Delta_n$$

for some constant $C > 0$. Similarly, we have that

$$F_{3,n} \xrightarrow{\mathbb{P}_{\theta^\dagger}} 2(\sigma^\dagger)^2. \tag{76}$$

From (74), (75) and (76), we immediately obtain the convergence (21).

Appendix F. Kalman filter for sub-class of (Hypo-II)

For simplicity, we write $x_i = (x_{S_1,i}, x_{S_2,i}, x_{R,i}) \in \mathbb{R}^N = \mathbb{R}^{N_{S_1}} \times \mathbb{R}^{N_{S_2}} \times \mathbb{R}^{N_R}$ for the state of scheme (22) at time t_i. Component $x_{S_1,i}$ is observable and $h_i = (x_{S_2,i}, x_{R,i}) \in \mathbb{R}^{N_H}$, $N_H = N_{S_1} + N_R$, is the hidden component, in agreement with applications. Thus, scheme (22) is now expressed as

$$x_{i+1} = b(\Delta_n, x_{S_1,i}, \theta) + A(\Delta_n, x_{S_1,i}, \theta) h_i + w(\Delta_n, \theta). \tag{77}$$

We set $\Sigma(\Delta_n, \theta) = \mathbb{E}\,[\, w(\Delta_n, \theta) w(\Delta_n, \theta)^\top]$ and assume that $h_0 | x_{S_1,0} \sim \mathscr{N}(m_0, Q_0)$ for some $m_0 \in \mathbb{R}^{N_H}$ and $Q_0 \in \mathbb{R}^{N_H \times N_H}$. Then, the filtering formula and the marginal likelihood are obtained as follows.

- *Filtering Recursion*: We have that

$$h_k | x_{S_1,0:k} \sim \mathscr{N}(m_k, Q_k), \quad 0 \le k \le n, \tag{78}$$

  with the filter mean m_k and covariance Q_k given as:

$$m_k = \mu_{H,k-1} + \Lambda_{HS_1,k-1}\, \Lambda^{-1}_{S_1S_1,k-1} \left(x_{S_1,k} - \mu_{S_1,k-1} \right);$$
$$Q_k = \Lambda_{HH,k-1} - \Lambda_{HS_1,k-1}\, \Lambda^{-1}_{S_1S_1,k-1} \Lambda_{S_1H,k-1},$$

  where $\mu_{H,k-1} \in \mathbb{R}^{N_H}$, $\mu_{S_1,k-1} \in \mathbb{R}^{N_{S_1}}$, $\Lambda_{S_1S_1,k-1} \in \mathbb{R}^{N_{S_1} \times N_{S_1}}$, $\Lambda_{S_1H,k-1} \in \mathbb{R}^{N_{S_1} \times N_H}$, $\Lambda_{HS_1,k-1} \in \mathbb{R}^{N_H \times N_{S_1}}$, $\Lambda_{HH,k-1} \in \mathbb{R}^{N_H \times N_H}$ are found via the following equations:

$$\mu_{k-1} = \begin{bmatrix} \mu_{S_1,k-1} \\ \mu_{H,k-1} \end{bmatrix} = b(\Delta_n, x_{S_1,k-1}, \theta) + A(\Delta_n, x_{S_1,k-1}, \theta) m_{k-1};$$
$$\Lambda_{k-1} = \begin{bmatrix} \Lambda_{S_1S_1,k-1} & \Lambda_{S_1H,k-1} \\ \Lambda_{HS_1,k-1} & \Lambda_{HH,k-1} \end{bmatrix} = \Sigma(\Delta_n, \theta) + A(\Delta_n, x_{S_1,k-1}, \theta)\, Q_{k-1}\, A(\Delta_n, x_{S_1,k-1}, \theta)^\top.$$

- *Marginal likelihood*: For a given initial distribution $p_\theta(x_{S_1,0})$, we have that

$$p_\theta(x_{S_1,0:n}) = p_\theta(x_{S_1,0}) \times \prod_{k=1}^{n} p_\theta(x_{S_1,k} | x_{S_1,0:k-1}), \tag{79}$$

  where $p_\theta(x_{S_1,k} | x_{S_1,0:k-1})$ is the density of $x_{S_1,k}$ given $x_{S_1,0:k-1}$ whose conditional distribution is given by:

$$x_{S_1,k} \,|\, x_{S_1,0:k-1} \sim \mathscr{N}(\mu_{S_1,k-1},\, \Lambda_{S_1S_1,k-1}).$$

F.1. Derivation of filter (78)

We assume that the filter in the previous time step is obtained as:

$$h_{k-1} | x_{S_1,0:k-1} \sim \mathscr{N}(m_{k-1}, Q_{k-1}). \tag{80}$$

It follows that

$$p_\theta(h_k | x_{S_1,0:k}) = \frac{p_\theta(x_k | x_{S_1,0:k-1})}{p_\theta(x_{S_1,k} | x_{S_1,0:k-1})},$$

and

$$p_\theta(x_k | x_{S_1,0:k-1}) = \int p_\theta(x_k, h_{k-1} | x_{S_1,0:k-1}) dh_{k-1} = \int p_\theta(x_k | x_{k-1}) p_\theta(h_{k-1} | x_{S_1,0:k-1}) dh_{k-1}. \tag{81}$$

From the definition of scheme (77), we have that

$$x_k | x_{k-1} \sim \mathscr{N}\big(b(\Delta_n, x_{S_1,k-1}, \theta) + A(\Delta_n, x_{S_1,k-1}, \theta) h_{k-1}, \Sigma(\Delta_n, \theta)\big). \tag{82}$$

From (80), (81) and (82), we obtain

$$x_k | x_{S_1,0:k-1} \sim \mathscr{N}\big(\mu_{k-1}, \Lambda_{k-1}\big), \tag{83}$$

where

$$\mu_{k-1} = b(\Delta_n, x_{S_1,k-1}, \theta) + A(\Delta_n, x_{S_1,k-1}, \theta) m_{k-1}, \qquad \Lambda_{k-1} = \Sigma(\Delta_n, \theta) + A(\Delta_n, x_{S_1,k-1}, \theta)\, Q_{k-1}\, A(\Delta_n, x_{S_1,k-1}, \theta)^\top.$$

Finally, applying the conditional Gaussian distribution formula, we obtain (78).

F.2. Derivation of the marginal likelihood (79)

From the marginal of the Gaussian distribution (83) for $x_k | x_{S_1,0:k-1}$, we immediately obtain:

$$x_{S_1,k} | x_{S_1,0:k-1} \sim \mathscr{N}\big(\mu_{S_1,k-1},\, \Lambda_{S_1S_1,k-1}\big),$$

where $\mu_{S_1,k-1} = \mathrm{proj}_{1,N_{S_1}}(\mu_{k-1})$, and $\Lambda_{S_1S_1,k-1} = \big[\Lambda^{ij}_{k-1}\big]_{1 \le i,j \le N_{S_1}}$.

References

[1] C. Ayaz, L. Tepper, F.N. Brünig, J. Kappler, J.O. Daldrop, R.R. Netz, Non-Markovian modeling of protein folding, Proc. Natl. Acad. Sci. USA 118 (31) (2021) e2023856118.
[2] V. Bally, D. Talay, The law of the Euler scheme for stochastic differential equations: I. Convergence rate of the distribution function, Probab. Theory Related Fields 104 (1996) 43–60.
[3] E. Buckwar, M. Tamborrino, I. Tubikanec, Spectral density-based and measure-preserving ABC for partially observed diffusion processes. An illustration on Hamiltonian SDEs, Stat. Comput. 30 (3) (2020) 627–648.
[4] T. Cass, Smooth densities for solutions to stochastic differential equations with jumps, Stoch. Process. Appl. 119 (5) (2009) 1416–1435.
[5] M. Ceriotti, G. Bussi, M. Parrinello, Colored-noise thermostats à la carte, J. Chem. Theory Comput. 6 (4) (2010) 1170–1180.
[6] N. Chen, Stochastic Methods for Modeling and Predicting Complex Dynamical Systems: Uncertainty Quantification, State Estimation, and Reduced-Order Models, Springer Nature, 2023.
[7] M. Coti Zelati, M. Hairer, A noise-induced transition in the Lorenz system, Comm. Math. Phys. 383 (2021) 2243–2274.
[8] S. Ditlevsen, A. Samson, Hypoelliptic diffusions: Filtering and inference from complete and partial observations, J. R. Stat. Soc., B: Stat. Methodol. 81 (2) (2019) 361–384.
[9] S. Ditlevsen, M. Tamborrino, I. Tubikanec, Network inference in a stochastic multi-population neural mass model via approximate Bayesian computation, 2023, arXiv preprint arXiv:2306.15787.
[10] R. Douc, E. Moulines, D. Stoffer, Nonlinear Time Series: Theory, Methods and Applications with R Examples, CRC Press, 2014.
[11] J. Dureau, K. Kalogeropoulos, M. Baguelin, Capturing the time-varying drivers of an epidemic using stochastic dynamical systems, Biostatistics 14 (3) (2013) 541–555.
[12] F. Ferretti, V. Chardes, T. Mora, A.M. Walczak, I. Giardina, Building general Langevin models from discrete datasets, Phys. Rev. X 10 (3) (2020) 031018.
[13] V. Genon-Catalot, J. Jacod, On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, Ann. l'IHP Probab. Stat. 29 (1) (1993) 119–151.
[14] A. Gloter, N. Yoshida, Adaptive and non-adaptive estimation for degenerate diffusion processes, 2020, arXiv preprint, arXiv:2002.10164.
[15] A. Gloter, N. Yoshida, Adaptive estimation for degenerate diffusion processes, Electron. J. Stat. 15 (1) (2021) 1424–1472.
[16] E. Gobet, C. Labart, Sharp estimates for the convergence of the density of the Euler scheme in small time, Electron. Commun. Probab. 13 (2008) 352–363.
[17] P. Hall, C.C. Heyde, Martingale Limit Theory and Its Application, Academic Press, 1980.
[18] Y. Iguchi, A. Beskos, M.M. Graham, Parameter estimation with increased precision for elliptic and hypo-elliptic diffusions, 2024, arxiv preprint, arXiv:2211.16384. In press in Bernoulli.
[19] Y. Iguchi, T. Yamada, Operator splitting around Euler-Maruyama scheme and high order discretization of heat kernels, ESAIM Math. Model. Numer. Anal. 55 (2021) S323–S367.
[20] S. Kalliadasis, S. Krumscheid, G.A. Pavliotis, A new framework for extracting coarse-grained models from time series with multiscale structure, J. Comput. Phys. 296 (2015) 314–328.
[21] M. Kessler, Estimation of an ergodic diffusion from discrete observations, Scand. J. Stat. 24 (2) (1997) 211–229.
[22] P.E. Kloeden, E. Platen, Numerical Solution of Stochastic Differential Equations, Springer, 1992.
[23] B. Leimkuhler, C. Matthews, Molecular dynamics: With deterministic and stochastic numerical methods, Interdiscip. Appl. Math. 39 (2015) 443.
[24] B. Leimkuhler, M. Sachs, Efficient numerical algorithms for the generalized Langevin equation, SIAM J. Sci. Comput. 44 (1) (2022) A364–A388.
[25] Z. Li, H.S. Lee, E. Darve, G.E. Karniadakis, Computing the non-Markovian coarse-grained interactions derived from the Mori–Zwanzig formalism in molecular systems: Application to polymer melts, J. Chem. Phys. 146 (1) (2017) 014104.
[26] B.G. Mitterwallner, C. Schreiber, J.O. Daldrop, J.O. Rädler, R.R. Netz, Non-Markovian data-driven modeling of single-cell motility, Phys. Rev. E 101 (3) (2020) 032408.
[27] H. Ness, L. Stella, C. Lorenz, L. Kantorovich, Applications of the generalized Langevin equation: Towards a realistic description of the baths, Phys. Rev. B 91 (1) (2015) 014301.
[28] D. Nualart, The Malliavin Calculus and Related Topics, vol. 1995, Springer, 2006.
[29] G.A. Pavliotis, Stochastic Processes and Applications: Diffusion Processes, the Fokker-Planck and Langevin Equations, vol. 60, Springer, 2014.
[30] P. Pigato, Tube estimates for diffusion processes under a weak Hörmander condition, Ann. Inst. H. Poincaré Probab. Statist. 54 (1) (2018) 299–342.
[31] P. Pilipovic, A. Samson, S. Ditlevsen, Parameter estimation in nonlinear multivariate stochastic differential equations based on splitting schemes, 2024, A preprint. hal-04457892.
[32] Y. Pokern, A.M. Stuart, P. Wiberg, Parameter estimation for partially observed hypoelliptic diffusions, J. R. Stat. Soc., B: Stat. Methodol. 71 (1) (2009) 49–73.
[33] A. Samson, M. Thieullen, A contrast estimator for completely or partially observed hypoelliptic diffusion, Stoch. Process. their Appl. 122 (7) (2012) 2521–2552.
[34] A. Spannaus, T. Papamarkou, S. Erwin, J.B. Christian, Inferring the spread of COVID-19: The role of time-varying reporting rate in epidemiological modelling, Sci. Rep. 12 (1) (2022) 1–12.
[35] H. Tsai, K. Chan, Testing for nonlinearity with partially observed time series, Biometrika 87 (4) (2000) 805–821.
[36] M. Uchida, N. Yoshida, Adaptive estimation of an ergodic diffusion process based on sampled data, Stoch. Process. Appl. 122 (8) (2012) 2885–2924.
[37] H. Vroylandt, L. Goudenège, P. Monmarché, F. Pietrucci, B. Rotenberg, Likelihood-based non-Markovian models from molecular dynamics, Proc. Natl. Acad. Sci. USA 119 (13) (2022) e2117586119.