DISCRETE-TIME STATISTICAL INFERENCE FOR MULTISCALE DIFFUSIONS[*]

SIRAGAN GAILUS[†] AND KONSTANTINOS SPILIOPOULOS[†]

Abstract. We study statistical inference for small-noise-perturbed multiscale dynamical systems under the assumption that we observe a single time series from the slow process only. We construct estimators for both averaging and homogenization regimes, based on an appropriate misspecified model motivated by a second-order stochastic Taylor expansion of the slow process with respect to a function of the time-scale separation parameter. In the case of a fixed number of observations, we establish consistency, asymptotic normality, and asymptotic statistical efficiency of a minimum contrast estimator (MCE), the limiting variance having been identified explicitly; we furthermore establish consistency and asymptotic normality of a simplified MCE, which is, however, not, in general, efficient. These results are then extended to the case of high-frequency observations under a condition restricting the rate at which the number of observations may grow vis-à-vis the separation of scales. Numerical simulations illustrate the theoretical results.

Key words. multiscale diffusions, slow fast, homogenization, averaging, parameter estimation

AMS subject classifications. 60F05, 60G25, 60J60, 62M02, 62M05

DOI. 10.1137/17M1147408

1. Introduction. Let us consider a family of $m + (d-m)$-dimensional processes $(X^\varepsilon, Y^{\varepsilon,})_T = \{(X_t^\varepsilon, Y_t^\varepsilon)\}_{0 \leq t \leq T}$ satisfying the stochastic differential equations (SDEs)

(1)

$$dX_t^\varepsilon = \frac{\epsilon}{\delta} b_\theta(X_t^\varepsilon, Y_t^\varepsilon)dt + c_\theta(X_t^\varepsilon, Y_t^\varepsilon)dt + \sqrt{\epsilon}\sigma(X_t^\varepsilon, Y_t^\varepsilon)dW_t,$$

$$dY_t^\varepsilon = \frac{\epsilon}{\delta^2} f(X_t^\varepsilon, Y_t^\varepsilon)dt + \frac{1}{\delta} g(X_t^\varepsilon, Y_t^\varepsilon)dt + \frac{\sqrt{\epsilon}}{\delta}\tau_1(X_t^\varepsilon, Y_t^\varepsilon)dW_t + \frac{\sqrt{\epsilon}}{\delta}\tau_2(X_t^\varepsilon, Y_t^\varepsilon)dB_t,$$

$$X_0^\varepsilon = x_0 \in \mathcal{X} = \mathbb{R}^m, Y_0^\varepsilon = y_0 \in \mathcal{Y} = \mathbb{R}^{d-m}.$$

Here, W and B are independent Wiener processes and $\varepsilon = (\epsilon, \delta) \in \mathbb{R}_+^2$ satisfy $\delta = \delta(\epsilon) \to 0$ as $\epsilon \to 0$. Note that δ is the time-scale separation parameter while ϵ dictates the size of the noise. The vector $\theta \in \Theta \subset \mathbb{R}^k$ represents an unknown parameter governing the drift coefficients b_θ and c_θ; the statistical problem considered herein is the estimation of θ based upon a time series sampled from a realization of the slow process X^ϵ.

Data from physical dynamical systems commonly exhibit multiple characteristic space or time scales. In addition, stochastic noise may be introduced to account for uncertainty or as an essential part of a particular modeling problem. Consequently, multiscale SDE models like (1) are widely employed in applied fields including physics, chemistry, and biology [4, 13, 32], neuroscience [14], meteorology [19], and econometrics and mathematical finance [7, 8, 9, 31] to describe stochastically perturbed dynamical systems with two or more different space or time scales.

[†]Department of Mathematics and Statistics, Boston University, Boston, MA 02215 (siragan@math.bu.edu, kspiliop@math.bu.edu).

The goal of this paper is to study the problem of statistical inference for the unknown parameter vector $\theta \in \Theta \subset \mathbb{R}^k$ based on a fixed discrete-time sample $\{x_{t_k}\}_{k=1}^n$ from an observation of the slow component X^ε. We consider this problem in the following two regimes:

1. ("∞ regime" or "homogenization regime") $\lim_{\epsilon \to 0} \frac{\epsilon}{\delta} = \infty$.
2. ("γ regime" or "averaging regime") $\lim_{\epsilon \to 0} \frac{\epsilon}{\delta} = \gamma \in (0, \infty)$.

In the ∞ regime, a standard "centering" condition (Condition 3) is imposed to regulate the asymptotically singular term $\frac{\epsilon}{\delta} b_\theta(X_t^\varepsilon, Y_t^\varepsilon) dt$ in the SDE for the slow component X^ε.

Statistical inference for diffusions without multiple scales (i.e., $\delta \equiv 1$) has been very well studied in the literature; see, for example, [3, 16, 18, 25]. In [3, 18, 25], $\epsilon \equiv \delta \equiv 1$ and the asymptotic behavior of the maximum likelihood estimator (MLE) based on continuous data is studied in the time horizon limit $T \to \infty$. This is directly analogous to the limit $n \to \infty$ in the classical setting of independent and identically distributed observations. The case of fixed time horizon T but $\epsilon \to 0$ has been studied in [16], likewise based on continuous data and without multiple scales. Estimation based on discrete data without multiple scales has also been addressed in the literature; see, for example, [12, 28, 29].

Maximum likelihood estimation from continuous data for multiscale models with noise of order one has been studied in [1, 2, 15, 20, 23]. The authors of [20, 23] prove that in the averaging regime, the MLE induced by the (nondeterministic) limit of the slow process X^ε in (1) with $\epsilon \equiv 1$ as $\delta \to 0$ is consistent under the assumption that coefficients are bounded and that the fast process Y^ε takes values in a torus. It is important to point out that the regime $\varepsilon \to 0$ which we study in this paper is different in that the diffusion $\sqrt{\epsilon}\sigma$ vanishes in the limit and, as described precisely by Theorem 1, X^ε converges to the solution of an ODE rather than an SDE; the (deterministic) limit does not induce a well-defined likelihood. Meanwhile, it is also shown in [20] that direct application of the principle of maximum likelihood with discretely sampled data via Euler–Maruyama approximation produces consistent estimates only if the data are first appropriately subsampled. Most closely related to the present work are [10, 27], wherein the authors prove consistency and asymptotic normality of the continuous-data MLE for special cases of (1).

Our focus in this paper is different. We address estimation from discrete data for multiscale diffusion models like (1). We assume that we are given only a discrete-time sample $\{x_{t_k}\}_{k=1}^n$ from a single observation of the slow process X^ε; we assume that no data are available from the fast process Y^ε. Motivated by a second-order stochastic Taylor expansion with respect to $\sqrt{\epsilon}$, we construct a minimum contrast estimator (hereinafter referred to as the MCE) based on an appropriate misspecified model. First, we prove that for any given fixed value of n, in either the averaging or the homogenization regime, this estimator is consistent and asymptotically normal as $\epsilon + \delta \to 0$ with a limiting variance $M(\theta; n)$ which we calculate explicitly. Knowing the limiting variance is important for statistical inference as it allows one to control the error, construct confidence intervals, and develop appropriate hypothesis tests. Going a step further, we show that $M(\theta; n)$ attains, in the limit as $n \to \infty$, the Cramér–Rao bound for the continuous-data estimation problem, which is to say that the estimator is asymptotically statistically efficient as first $\epsilon + \delta \to 0$ and then $n \to \infty$. Second, we study a simplified MCE (hereinafter referred to as the SMCE) that can be considerably easier to apply and presents improved robustness in numerical studies. We show that the simplified estimator is also consistent and asymptotically normal, although it is not in general efficient. Third, we study the behavior of the two estimators in the *joint*

limit $\epsilon + \delta \to 0$ *and* $n \to \infty$ (the high-frequency regime), showing that consistency and asymptotic normality (and, for the MCE, asymptotic efficiency) are retained provided that the sampling interval $\Delta := T/n$ does not decrease too quickly relative to ϵ (or, equivalently, depending on the regime, to δ). In particular, we prove in the high-frequency regime (a) consistency of both MCE and SMCE under the assumption that $\epsilon = o(\Delta)$, and (b) asymptotic normality of the MCE under the stronger assumption $\epsilon = o(\Delta^2)$ but asymptotic normality of SMCE under the assumption $\epsilon = o(\Delta)$ only. It is clear that these conditions can be written also in terms of δ and Δ.

To the best of our knowledge, this is the first paper on discrete-time estimation for multiscale models to describe estimators demonstrated to be consistent, asymptotically normal, and asymptotically statistically efficient. The limiting variance of the estimators is calculated explicitly. For high-frequency observations, we require that Δ not decrease too quickly relative to ϵ; this is reminiscent of the subsampling prescribed in [2, 20], although our case is different in that we take T to be fixed, and so also the relationship between $\Delta := T/n$ and n is fixed. Despite our best efforts, we have not managed to relax these assumptions; nevertheless, numerical simulations suggest that the estimators remain well behaved, even when ϵ is not of lower order than Δ (see Tables 5 and 6 and Figures 9–12 in section 7). We emphasize that our estimators can be applied without precise knowledge of δ (which is tricky to estimate in practice) or ϵ.

Let us now briefly discuss the approach that we will take. The main idea is to establish a second-order stochastic Taylor expansion of X^ε of the form

$$X^\varepsilon = \bar{X} + \sqrt{\epsilon}\varphi^\varepsilon,$$

where $\bar{X}$ is the deterministic law-of-large-numbers limit of X^ε and φ^ε converges in distribution to a Gaussian process as $\epsilon \to 0$. This representation motivates a misspecified model in which the conditional distribution of $X^\varepsilon_{t_k}$ given $X^\varepsilon_{t_{k-1}} = x_{t_{k-1}}$ is approximated by a Gaussian random variable having a specified mean and variance. The principle of maximum likelihood applied to this discrete, approximate process leads us to a certain contrast function, the minimizer of which we take to be our estimator. Naturally, appropriate identifiability conditions must be assumed; these conditions are, however, typically satisfied provided that the coefficients in the original model are sufficiently regular. We mention here that we have restricted the dependence of (1) on the parameter θ to the coefficients b and c for purposes of presentation only; as inspection of the proofs will make apparent, the results of the paper continue to hold unchanged if one allows the other coefficients $\sigma, f, g, \tau_1, \tau_2$ also to depend on θ.

The rest of this paper is organized as follows. Section 2 presents the main assumptions of the paper and introduces helpful notation. Section 3 develops a second-order stochastic Taylor expansion of X^ε; the representation thus obtained motivates the methods of statistical inference presented in later sections. Section 4 develops the proposed estimator assuming a fixed number n of data points. We prove that the estimator is consistent and asymptotically normal as $\epsilon + \delta \to 0$, and asymptotically statistically efficient in the sense that the limiting variance, as a function of the number of observations n, attains the Cramér–Rao bound as $n \to \infty$. Section 5 studies a simplification of the estimator of section 4 that offers certain advantages in practice at the cost of an increase in limiting variance. Section 6 studies the joint limit $\epsilon + \delta \to 0$ and $n \to \infty$, concluding that the proof of consistency goes through provided that one has $\epsilon \cdot n \to 0$, which is to say that the squared diffusion vanishes more quickly than does the sampling interval $\Delta := T/n$. Similarly, asymptotic normality holds for the SMCE provided that $\epsilon = o(\Delta)$ and for the MCE if $\epsilon = o(\Delta^2)$. Section 7 presents

the results of numerical simulations to illustrate the theoretical results; simulations are described also that deliberately violate the relations $\epsilon = o(\Delta)$ and $\epsilon = o(\Delta^2)$ that we impose for the high-frequency theory, in order to substantiate our conjecture that it may yet be possible to weaken these assumptions. Section 8 summarizes our conclusions and discusses future directions of the research. Finally, an appendix collects technical estimates used throughout the paper.

2. Preliminaries and assumptions. Let us begin with a discussion of the main assumptions that we carry throughout the paper. We work with a canonical probability space $(\Omega, \mathcal{F}, P)$ equipped with a filtration $\{\mathcal{F}_t\}_{0 \le t \le T}$ satisfying the usual conditions (namely, $\{\mathcal{F}_t\}_{0 \le t \le T}$ is right continuous and $\mathcal{F}_0$ contains all P-negligible sets). Recall from (1) that $\mathcal{X} = \mathbb{R}^m$ and $\mathcal{Y} = \mathbb{R}^{d-m}$ are, respectively, the state spaces of the slow and fast components of the dynamics.

To avoid ambiguity, we will always write $|\cdot|$ for the Frobenius (Euclidean) norm, and $||\cdot||$ for the operator norm of a matrix.

To guarantee that (1) is well posed and that our limit results are valid, we impose the following regularity and growth conditions.

CONDITION 1 (regularity of coefficients).

Conditions on c_θ:
1. $\exists K > 0, q > 0, r \in [0,1) \ \forall \theta \in \Theta, |c_\theta(x,y)| \le K(1 + |x|^r)(1 + |y|^q)$.
2. $\exists K > 0, q > 0 \ \forall \theta \in \Theta, |\nabla_x c_\theta(x,y)| + |\nabla_x \nabla_x c_\theta(x,y)| \le K(1 + |y|^q)$.
3. $\forall \theta \in \Theta$, c_θ has two continuous derivatives in x, Hölder continuous in y uniformly in x.
4. $\forall \theta \in \Theta, \nabla_y \nabla_y c_\theta(x,y)$ is jointly continuous in x and y.
5. $c_\theta(x,y)$ has two locally bounded derivatives in θ with at most polynomial growth in x and y.

Conditions on σ:
1. $\forall N > 0, \exists C(N) \ \forall x_1, x_2 \in \mathcal{X} \ \forall y \in \mathcal{Y}$ with $|y| \le N, |\sigma(x_1,y) - \sigma(x_2,y)| \le C(N)|x_1 - x_2|$.
2. $\exists K > 0, q > 0, |\sigma(x,y)| \le K(1 + |x|^{1/2})(1 + |y|^q)$.
3. $\sigma\sigma^T$ is uniformly nondegenerate.

Conditions on f, τ_1, τ_2:
1. $f, \tau_1\tau_1^T$, and $\tau_2\tau_2^T$ are twice differentiable in x and y, the first and second derivatives in x being bounded, and all partial derivatives up to second order being Hölder continuous in y uniformly in x.
2. $\tau_2\tau_2^T$ is uniformly nondegenerate.

Conditions on b_θ, g:
1. b_θ satisfies the same smoothness conditions as c_θ.
2. In the ∞ regime, b_θ and its derivatives are bounded uniformly in the first variable by polynomials in the second variable.
3. In the ∞ regime, g satisfies the same conditions as c_θ; in the γ regime, g satsifies the same conditions as f.

In the limit of infinite-scale separation, the slow process appears from the perspective of the fast to become "frozen." To guarantee that the fast process has an invariant distribution when the slow process is frozen, we impose the following recurrence condition:

CONDITION 2 (recurrence condition).
1. In the ∞ regime, $\lim_{|y| \to \infty} \sup_{x \in \mathcal{X}}(f(x,y) \cdot y) = -\infty$.
2. In the γ regime, $\lim_{|y| \to \infty} \sup_{x \in \mathcal{X}}((\gamma f + g)(x,y) \cdot y) = -\infty$.

The conditions on f, τ_1, τ_2 in Conditions 1 and 2 guarantee that for each fixed $x \in \mathcal{X}$ one has on $\mathcal{Y}$, in the ∞ regime, a unique invariant measure $\mu_{\infty,x}$ associated with the operator

$$\mathcal{L}_{\infty,x} := f(x, \cdot) \cdot \nabla_y + \frac{1}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, \cdot) : \nabla_y^2,$$

and in the γ regime, a unique invariant measure $\mu_{\gamma,x}$ associated with the operator

$$\mathcal{L}_{\gamma,x} := (\gamma f + g)(x, \cdot) \cdot \nabla_y + \frac{\gamma}{2}(\tau_1 \tau_1^T + \tau_2 \tau_2^T)(x, \cdot) : \nabla_y^2.$$

For existence of an invariant measure, the interested reader may see, for example, [30]; uniqueness is a consequence of nondegenerate diffusion, as for example in [22].

In the ∞ regime, a standard "centering" condition is imposed to regulate the asymptotically singular term $\frac{\epsilon}{\delta} b_\theta(X_t^\varepsilon, Y_t^\varepsilon) dt$ in order that we may obtain a homogenization limit.

CONDITION 3 (centering condition for the ∞ regime). *For each fixed $x \in \mathcal{X}$,* $\int_{\mathcal{Y}} b(x, y) \mu_{\infty,x}(dy) = 0.$

We also need a condition on the relative rates at which ϵ and δ vanish in order to derive a suitable second-order approximation of the slow component X^ε.

CONDITION 4.
 1. *In the ∞ regime, $\frac{\epsilon^{3/2}}{\delta} \to \ell_\infty$ as $\epsilon \to 0$ for some $\ell_\infty \in (0, \infty]$.*
 2. *In the γ regime, $\frac{\sqrt{\epsilon}}{\frac{\epsilon}{\delta} - \gamma} \to \ell_\gamma$ as $\epsilon \to 0$ for some $\ell_\gamma \in (0, \infty]$.*

Let us conclude this section with an introduction of notational conventions. First, an asterisk in place of the regime signifier (writing $\mathcal{L}_{*,x}$, $\mu_{*,x}$, etc.) will be used when convenient if a statement is to be understood for both regimes. Next, we will in many instances wish to integrate functions of two variables x and y in the second variable over the invariant measures $\mu_{*,x}$ to obtain "averaged" functions of x only. In such cases a bar will distinguish an averaged function from the original function; that is, for a generic function h of x and y,

$$\bar{h}(x) := \int_{\mathcal{Y}} h(x, y) \mu_{*,x}(dy).$$

This bar notation should not be confused with $\bar{X}$, which is the first-order limit of the stochastic process X^ε, nor with $\bar{\theta}$, which is an estimator. Finally, we will denote by $A : B$ the Frobenius inner product $\Sigma_{i,j}[a_{i,j} \cdot b_{i,j}]$ of matrices $A = (a_{i,j})$ and $B = (b_{i,j})$.

3. Asymptotic behavior of X^ε. In this section, we develop a stochastic Taylor expansion of X^ε, establishing in particular a representation

(2) $$X^\varepsilon = \bar{X}_* + \sqrt{\epsilon} \varphi_*^\varepsilon$$

in which $\bar{X}_*$ is deterministic and φ_*^ε converges in distribution to a Gaussian process as $\epsilon \to 0$. This representation is key in that it allows us to define estimators in terms

of an appropriate misspecified model and prove consistency and asymptotic normality (see sections 4, 5, and 6). We suppress the parameter θ as the results of this section hold independently of (and indeed uniformly over) $\theta \in \Theta$.

Intuitively speaking, as $\delta \to 0$, the fast dynamics accelerate relative to the slow and Y_t^ε tends to its invariant distribution at each frozen value X_t^ε. At the same time, the driving noise $\sqrt{\epsilon}\sigma dW$ of the slow process tends to 0. It is therefore reasonable to anticipate that, in the limit, X^ε should hew closely to the solution $\bar{X}_*$ of an ODE involving coefficients averaged over the fast dynamics.

Recall that in the ∞ regime, $\frac{\epsilon}{\delta} \to \infty$; to describe the asymptotic behavior of X^ε in this regime, we need to find the limiting contribution of the asymptotically singular term $\frac{\epsilon}{\delta} b(X_t^\varepsilon, Y_t^\varepsilon) dt$. It turns out that under Condition 3, the limiting contribution may be captured in terms of the solution of a certain Poisson equation. By Theorem 3 in [22], there is a unique solution χ in the class of functions that grows at most polynomially in $|y|$ as $y \to \infty$ of the equation

$$(3) \qquad \mathcal{L}_{\infty,x}\chi(x,y) = -b(x,y), \quad \int_{\mathcal{Y}} \chi(x,y)\mu_{\infty,x}(dy) = 0.$$

Theorem 1 establishes that the averaged coefficient in the ODE for $\bar{X}_\infty$ is $\bar{\lambda}_\infty$, where

$$(4) \qquad \lambda_\infty(x,y) := (\nabla_y \chi \cdot g + c)(x,y).$$

In the γ regime, there is no singular coefficient with which to contend, and the corresponding averaged coefficient is $\bar{\lambda}_\gamma$, where

$$(5) \qquad \lambda_\gamma(x,y) := (\gamma b + c)(x,y).$$

Let us state this law-of-large-numbers approximation precisely in the form of a theorem. This gives a first-order approximation to X^ε when ϵ, δ are small.

THEOREM 1. *Assume Conditions 1 and 2 and, in the ∞ regime, Condition 3; let $*$ denote the regime. For any initial condition $(x_0, y_0) \in \mathcal{X} \times \mathcal{Y}$ and $0 < p < \infty$, there is a constant $\tilde{K}$ such that for ϵ sufficiently small,*

$$E \sup_{0 \le t \le T} |X_t^\varepsilon - \bar{X}_{*,t}|^p \le \tilde{K} \cdot \epsilon^{p/2},$$

where $\bar{X}_*$ is the (deterministic) solution of the integral equation

$$\bar{X}_{*,t} := x_0 + \int_0^t \bar{\lambda}_*(\bar{X}_{*,s}) ds,$$

where $\bar{\lambda}_*$ is obtained, depending on the regime, by averaging (4) or (5) over the invariant measures $\mu_{*,x}$.

Theorem 1 extends Theorem 1 in [10]. The proof relies on Lemma 7 in the appendix; we omit the details, as given Lemma 7, the argument follows nearly verbatim the proof of Theorem 1 in [10].

The essential content of Theorem 1 is that X^ε tends to a deterministic first-order limit. The minimum contrast estimators that we study in this paper exploit

the structure of the random fluctuations of X^ε about this limit; to do this, we must obtain a description of the asymptotic behavior to higher order.

This, in turn, turns on quantifying more precisely the difference between the true drift $\frac{\epsilon}{\delta}b(X_t^\varepsilon, Y_t^\varepsilon) + c(X_t^\varepsilon, Y_t^\varepsilon)$ and the approximate drift $\bar{\lambda}_*(\bar{X}_t)$. The limiting coefficient λ_* plays the role of intermediary. By Theorem 3 in [22], there is a unique solution Φ_* in the class of functions that grow at most polynomially in $|y|$ as $y \to \infty$ of the equation

$$(6) \qquad \mathcal{L}_{*,x}\Phi_*(x,y) = -\left(\lambda_* - \bar{\lambda}_*\right)(x,y), \quad \int_{\mathcal{Y}} \Phi_*(x,y)\mu_{*,x}(dy) = 0.$$

The function Φ will feature in our description of the fluctuations.

We need to define one more quantity before stating the next result. For any $0 \le s \le t \le T$, let $Z_*(t,s)$ denote the matrix-valued solution to the equation

$$(7) \qquad \frac{dZ_*(t,s)}{dt} = (\nabla_x\bar{\lambda}_*)(\bar{X}_{*,t})Z_*(t,s), \quad Z_*(s,s) = 1_m,$$

where 1_m denotes the $m \times m$ identity matrix. By Proposition 2.14 in [5], the continuity of $t \mapsto (\nabla_x\bar{\lambda}_*)(\bar{X}_{*,t})$ on $(0,T]$ guarantees the semigroup relations $Z_*(t,s) = Z_*(t,\rho)Z_*(\rho,s)$ and the invertibility of $Z_*(t,s)$.

THEOREM 2. *Assume Conditions 1, 2, and 4, and, in the ∞ regime, Condition 3.*

In the ∞ regime, we have a representation

$$(8) \qquad \frac{1}{\sqrt{\epsilon}}(X_t^\varepsilon - \bar{X}_{\infty,t}) = \frac{1}{\ell_\infty}\int_0^t Z_\infty(t,s)\overline{\nabla_y\Phi_\infty \cdot g}(\bar{X}_{\infty,s})ds$$

$$+ \int_0^t Z_\infty(t,s)(\sigma + \nabla_y\chi \cdot \tau_1)(X_s^\varepsilon, Y_s^\varepsilon)dW_s$$

$$+ \int_0^t Z_\infty(t,s)\nabla_y\chi \cdot \tau_2(X_s^\varepsilon, Y_s^\varepsilon)dB_s + \tilde{\mathcal{R}}_{\infty,t}^\varepsilon$$

such that for any $\eta > 0$, we have $\lim_{\epsilon \to 0} P(\sup_{0 \le t \le T}|\tilde{\mathcal{R}}_{\infty,t}^\varepsilon| > \eta) = 0$.

In the γ regime, we have a representation

$$(9) \qquad \frac{1}{\sqrt{\epsilon}}(X_t^\varepsilon - \bar{X}_{\gamma,t}) = \frac{1}{\ell_\gamma \cdot \gamma}\int_0^t Z_\gamma(t,s)\overline{\gamma b - \nabla_y\Phi_\gamma \cdot g}(\bar{X}_{\gamma,s})ds$$

$$+ \int_0^t Z_\gamma(t,s)(\sigma + \nabla_y\Phi_\gamma \cdot \tau_1)(X_s^\varepsilon, Y_s^\varepsilon)dW_s$$

$$+ \int_0^t Z_\gamma(t,s)\nabla_y\Phi_\gamma \cdot \tau_2(X_s^\varepsilon, Y_s^\varepsilon)dB_s + \tilde{\mathcal{R}}_{\gamma,t}^\varepsilon$$

such that for any $\eta > 0$, we have $\lim_{\epsilon \to 0} P(\sup_{0 \le t \le T}|\tilde{\mathcal{R}}_{\gamma,t}^\varepsilon| > \eta) = 0$.

Remark 1. Before proceeding to the proof, we mention that a similar result, albeit one insufficient for our purposes, is established in [26]. It is proven there that the fluctuations process $\eta_t^\varepsilon := \frac{X_t^\varepsilon - \bar{X}_t}{\sqrt{\epsilon}}$ converges weakly in the space of continuous functions $\mathcal{C}\left([0,T], \mathbb{R}^m\right)$ to the solution of a certain Ornstein–Uhlenbeck-type process η_t (with different dynamics for the different regimes). Unfortunately, this result is of limited use in the statistical setting because the limit is in distribution, which means that prelimit and limit do not live necessarily, as such, on the same space; hence the need for Theorem 2, which establishes a representation in path space.

Proof of Theorem 2. Let us suppress the regime subscript and suppose at first that we are in the ∞ regime. We write

$$\frac{1}{\sqrt{\epsilon}}(X_t^\varepsilon - \bar{X}_t) = \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where

$$\mathrm{I} := \frac{1}{\sqrt{\epsilon}} \int_0^t \left(\bar{\lambda}(X_s^\varepsilon) - \bar{\lambda}(\bar{X}_s) \right) ds,$$

$$\mathrm{II} := \frac{1}{\sqrt{\epsilon}} \int_0^t \left(\lambda(X_s^\varepsilon, Y_s^\varepsilon) - \bar{\lambda}(X_s^\varepsilon) \right) ds,$$

$$\mathrm{III} := \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s + \frac{1}{\sqrt{\epsilon}} \int_0^t \left(\frac{\epsilon}{\delta} b - \nabla_y \chi \cdot g \right)(X_s^\varepsilon, Y_s^\varepsilon) ds.$$

We begin with an approximation of each part.

$$\mathrm{I} = \int_0^t (\nabla_x \bar{\lambda})(\bar{X}_s) \cdot \frac{1}{\sqrt{\epsilon}}(X_s^\varepsilon - \bar{X}_s) ds$$

$$+ \int_0^t \left[(\nabla_x \bar{\lambda})(X_s^{\varepsilon,\dagger}) - (\nabla_x \bar{\lambda})(\bar{X}_s) \right] \cdot \frac{1}{\sqrt{\epsilon}}(X_s^\varepsilon - \bar{X}_s) ds$$

$$=: \int_0^t (\nabla_x \bar{\lambda})(\bar{X}_s) \cdot \frac{1}{\sqrt{\epsilon}}(X_s^\varepsilon - \bar{X}_s) ds + \mathcal{R}_{\mathrm{I},t}^\varepsilon,$$

where $X_s^{\epsilon,\dagger}$ is an appropriately chosen point on the segment connecting X_s^ε with $\bar{X}_s$.

$$\mathcal{R}_{\mathrm{I},t}^\varepsilon := \int_0^t \left[(\nabla_x \bar{\lambda})(X_s^{\varepsilon,\dagger}) - (\nabla_x \bar{\lambda})(\bar{X}_s) \right] \cdot \frac{1}{\sqrt{\epsilon}}(X_s^\varepsilon - \bar{X}_s) ds$$

vanishes in probability uniformly in t as $\epsilon \to 0$; to see this, note first that $\int_0^t \left[(\nabla_x \bar{\lambda})(X_s^{\epsilon,\dagger}) - (\nabla_x \bar{\lambda})(\bar{X}_s) \right] ds$ vanishes in probability by compactness of $[0,T]$, continuity of $\nabla_x \bar{\lambda}$, and Theorem 1, and second that $\sup_{0 \leq t \leq T} |\frac{1}{\sqrt{\epsilon}}(X_t^\varepsilon - \bar{X}_t)|$ is bounded in probability by Theorem 1.

Letting Φ be as in (6), applying the Itô formula to $\Phi(x,y)$ with $(x,y) = (X_t^\varepsilon, Y_t^\varepsilon)$, and rearranging terms, we have

(10)

$$
\begin{aligned}
\mathrm{II} ={}& \frac{\delta}{\epsilon^{3/2}} \int_0^t \nabla_y \Phi \cdot g(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \frac{\delta}{\sqrt{\epsilon}} \int_0^t \nabla_x \Phi \cdot b(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\delta^2}{\epsilon^{3/2}} \int_0^t \nabla_x \Phi \cdot c(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \frac{\delta^2}{\sqrt{\epsilon}} \int_0^t \sigma\sigma^T : \nabla_x^2 \Phi(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \frac{\delta}{\sqrt{\epsilon}} \int_0^t \sigma\tau_1^T : \nabla_y \nabla_x \Phi(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\delta^2}{\epsilon} \int_0^t \nabla_x \Phi \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
& + \frac{\delta}{\epsilon} \int_0^t (\nabla_y \Phi \cdot \tau_1, \nabla_y \Phi \cdot \tau_2)(X_s^\varepsilon, Y_s^\varepsilon) d(W,B)_s + \frac{\delta^2}{\epsilon^{3/2}} (\Phi(X_t^\varepsilon, Y_t^\varepsilon) - \Phi(x_0, y_0)) \\
={}& \frac{1}{\ell} \int_0^t \overline{\nabla_y \Phi \cdot g}(\bar{X}_s) ds + \frac{1}{\ell} \int_0^t \left[\nabla_y \Phi \cdot g(X_s^\varepsilon, Y_s^\varepsilon) - \overline{\nabla_y \Phi \cdot g}(\bar{X}_s) \right] ds \\
& + \left(\frac{\delta}{\epsilon^{3/2}} - \frac{1}{\ell} \right) \int_0^t \nabla_y \Phi \cdot g(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \frac{\delta}{\sqrt{\epsilon}} \int_0^t \nabla_x \Phi \cdot b(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\delta^2}{\epsilon^{3/2}} \int_0^t \nabla_x \Phi \cdot c(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \frac{\delta^2}{\sqrt{\epsilon}} \int_0^t \sigma\sigma^T : \nabla_x^2 \Phi(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \frac{\delta}{\sqrt{\epsilon}} \int_0^t \sigma\tau_1^T : \nabla_y \nabla_x \Phi(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\delta^2}{\epsilon} \int_0^t \nabla_x \Phi \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
& + \frac{\delta}{\epsilon} \int_0^t (\nabla_y \Phi \cdot \tau_1, \nabla_y \Phi \cdot \tau_2)(X_s^\varepsilon, Y_s^\varepsilon) d(W,B)_s \\
& - (\delta/\epsilon)(\delta/\sqrt{\epsilon})(\Phi(X_t^\varepsilon, Y_t^\varepsilon) - \Phi(x_0, y_0)) \\
=:{}& \frac{1}{\ell} \int_0^t \overline{\nabla_y \Phi \cdot g}(\bar{X}_s) ds + \mathcal{R}_{\mathrm{II},t}^\varepsilon.
\end{aligned}
$$

$\mathcal{R}_{\mathrm{II}}^\varepsilon$ vanishes in probability uniformly in t as $\epsilon \to 0$; to see this, note first that $\frac{1}{\ell} \int_0^t \left[\nabla_y \Phi \cdot g(X_s^\varepsilon, Y_s^\varepsilon) - \overline{\nabla_y \Phi \cdot g}(\bar{X}_s) \right] ds$ vanishes in probability by Lemma 7, second that $(\delta/\sqrt{\epsilon})(\Phi(X_t^\varepsilon, Y_t^\varepsilon) - \Phi(x_0, y_0))$ vanishes in probability by the argument of Corollary 1 in [21], and third that the remaining integrals are bounded in probability by Lemma 5 while their prefactors vanish.

Letting χ be as in (3), applying the Itô formula to $\chi(x, y)$ with $(x, y) = (X_t^\varepsilon, Y_t^\varepsilon)$, and rearranging terms, we have

$$
\begin{aligned}
\text{III} = & \int_0^t (\sigma + \nabla_y \chi \cdot \tau_1)(X_s^\varepsilon, Y_s^\varepsilon) dW_s + \int_0^t \nabla_y \chi \cdot \tau_2(X_s^\varepsilon, Y_s^\varepsilon) dB_s \\
& + \sqrt{\epsilon} \int_0^t \nabla_x \chi \cdot b(X_s^\varepsilon, Y_s^\varepsilon) ds + \frac{\delta}{\sqrt{\epsilon}} \int_0^t \nabla_x \chi \cdot c(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \sqrt{\epsilon} \delta \int_0^t \sigma \sigma^T : \nabla_x^2 \chi(X_s^\varepsilon, Y_s^\varepsilon) ds \\
& + \sqrt{\epsilon} \int_0^t \sigma \tau_1^T : \nabla_y \nabla_x \chi(X_s^\varepsilon, Y_s^\varepsilon) ds + \delta \int_0^t \nabla_x \chi \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s \\
& - (\delta/\sqrt{\epsilon}) \left(\chi(X_t^\varepsilon, Y_t^\varepsilon) - \chi(x_0, y_0) \right) \\
=: & \int_0^t (\sigma + \nabla_y \chi \cdot \tau_1)(X_s^\varepsilon, Y_s^\varepsilon) dW_s + \int_0^t \nabla_y \chi \cdot \tau_2(X_s^\varepsilon, Y_s^\varepsilon) dB_s + \mathcal{R}_{\text{III},t}^\varepsilon.
\end{aligned}
$$

$\mathcal{R}_{\text{III}}^\varepsilon$ vanishes in probability uniformly in t as $\epsilon \to 0$; to see this, note first that $(\delta/\sqrt{\epsilon}) \left(\chi(X_t^\varepsilon, Y_t^\varepsilon) - \chi(x_0, y_0) \right)$ vanishes in probability by the argument of Corollary 1 in [21] and, second, that the remaining integrals are bounded in probability by Lemma 5 while their prefactors vanish.

Putting everything together and setting $\mathcal{R}^\varepsilon := \mathcal{R}_I^\varepsilon + \mathcal{R}_{\text{II}}^\varepsilon + \mathcal{R}_{\text{III}}^\varepsilon$, we have

$$
\begin{aligned}
\frac{1}{\sqrt{\epsilon}}(X_t^\varepsilon - \bar{X}_t) = & \int_0^t (\nabla_x \bar{\lambda})(\bar{X}_s) \cdot \frac{1}{\sqrt{\epsilon}}(X_s^\varepsilon - \bar{X}_s) ds + \frac{1}{\ell} \int_0^t \overline{\nabla_y \Phi \cdot g}(\bar{X}_s) ds \\
& + \int_0^t (\sigma + \nabla_y \chi \cdot \tau_1)(X_s^\varepsilon, Y_s^\varepsilon) dW_s + \int_0^t \nabla_y \chi \cdot \tau_2(X_s^\varepsilon, Y_s^\varepsilon) dB_s + \mathcal{R}_t^\varepsilon.
\end{aligned}
$$

The desired representation follows; $\tilde{\mathcal{R}}^\varepsilon$ vanishes in probability by the same arguments as did $\mathcal{R}^\varepsilon$.

The proof for the γ regime is similar and is therefore omitted. This concludes the proof of the theorem. $\square$

Remark 2. It is not difficult to see that if, for example, $\nabla_x \bar{\lambda}$ is globally Lipschitz, then in fact $\mathcal{R}^\varepsilon$ and $\tilde{\mathcal{R}}^\varepsilon$ vanish in $\mathcal{L}^p$. This stronger convergence is, however, not necessary for our purposes.

4. Statistical inference for a fixed number of observations. Let us now use Theorem 2 to motivate the statistical estimators for the unknown parameter vector θ. Suppose that the drift coefficients in the model are known functions of an unknown parameter θ which we wish to estimate using a discrete-time sample $\{x_{t_k}\}_{k=1}^n$ from the slow process; here $n > 0$ is a fixed positive integer and $\{t_k\}_{k=1}^n \subset (0, T]$ is an increasing sequence of positive times. To simplify the presentation, let us assume a uniform sampling interval $\Delta := T/n$ and adopt the convention that $t_0 = 0$; hence $t_k = k\Delta$ for $k = 0, 1, \ldots, n$.

To motivate our estimator, recall the representation of X^ε established in Theorem 2. Set, in the ∞ regime,

$$
J_\infty^\theta(x, y) := \frac{1}{\ell_\infty}(\nabla \Phi_\infty \cdot g)^\theta(x, y),
$$
$$
q_\infty^\theta(x, y) := [(\sigma + \nabla_y \chi \cdot \tau_1)(\sigma + \nabla_y \chi \cdot \tau_1)^T + (\nabla_y \chi \cdot \tau_2)(\nabla_y \chi \cdot \tau_2)^T]^\theta(x, y)
$$

and, in the γ regime,

$$J_\gamma^\theta(x, y) := \frac{1}{\ell_\gamma \cdot \gamma}(\gamma b - \nabla_y \Phi_\gamma \cdot g)^\theta(x, y),$$

$$q_\gamma^\theta(x, y) := [(\sigma + \nabla_y \Phi_\gamma \cdot \tau_1)(\sigma + \nabla_y \Phi_\gamma \cdot \tau_1)^T + (\nabla_y \Phi_\gamma \cdot \tau_2)(\nabla_y \Phi_\gamma \cdot \tau_2)^T]^\theta(x, y).$$

Omitting the regime subscript and denoting by $Z^\theta(t, s)$ the θ-dependent solution to (7), (9) and (10) suggest heuristically that

$$(11) \quad X_t^{\varepsilon,\theta} - \bar{X}_t^\theta \approx \sqrt{\epsilon} \int_0^t Z^\theta(t, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds + \sqrt{\epsilon} \int_0^t Z^\theta(t, s) \cdot (\bar{q}^\theta)^{1/2}(\bar{X}_s^\theta) d\tilde{W}_s,$$

where $\tilde{W}$ is a Wiener process. We emphasize that this approximation is not mathematically correct. Apart from ignoring the remainder term $\tilde{\mathcal{R}}^\varepsilon$, we have replaced the Itô integral terms with a new Itô integral matching the limiting variance. Nevertheless, proceeding from (11),

$$\left[X_{t_k}^{\varepsilon,\theta} - \bar{X}_{t_k}^\theta\right] - Z^\theta(t_k, t_{k-1}) \cdot \left[X_{t_{k-1}}^{\varepsilon,\theta} - \bar{X}_{t_{k-1}}^\theta\right]$$

$$\approx \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds + \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot (\bar{q}^\theta)^{1/2}(\bar{X}_s^\theta) d\tilde{W}_s,$$

or what is the same upon rearranging terms,

$$X_{t_k}^{\varepsilon,\theta} \approx \bar{X}_{t_k}^\theta + Z^\theta(t_k, t_{k-1}) \cdot \left[X_{t_{k-1}}^{\varepsilon,\theta} - \bar{X}_{t_{k-1}}^\theta\right] + \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds$$

$$+ \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot (\bar{q}^\theta)^{1/2}(\bar{X}_s^\theta) d\tilde{W}_s.$$

In this way, it would appear that for each k, the conditional distribution of $X_{t_k}^\varepsilon$ given $X_{t_{k-1}}^\varepsilon = x_{t_{k-1}}$ is approximated by a Gaussian distribution with mean

$$\bar{X}_{t_k}^\theta + Z^\theta(t_k, t_{k-1}) \cdot \left[x_{t_{k-1}} - \bar{X}_{t_{k-1}}^\theta\right] + \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds$$

and variance $\epsilon \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{q}^\theta(\bar{X}_s^\theta) \cdot (Z^\theta)^T(t_k, s) ds$.

This heuristic Gaussian approximation to the increments may be regarded as a misspecified model, the principle of maximum likelihood applied to which suggests the MCE

$$(12) \quad \bar{\theta}^\varepsilon(\{x_{t_k}\}_{k=1}^n) := \arg\min_{\theta \in \bar{\Theta}} U^\varepsilon(\theta; \{x_{t_k}\}_{k=1}^n)$$

with contrast function

$$(13) \quad U^\varepsilon(\theta; \{x_{t_k}\}_{k=1}^n) := \Sigma_{k=1}^n \left[(F_k^\varepsilon)^T(\theta; \{x_{t_k}\}_{k=1}^n) \cdot Q_k^{-1}(\theta) \cdot F_k^\varepsilon(\theta; \{x_{t_k}\}_{k=1}^n)\right],$$

where

$$(14) \quad F_k^\varepsilon(\theta; \{x_{t_k}\}_{k=1}^n) := \left[[x_{t_k} - \bar{X}_{t_k}^\theta] - Z^\theta(t_k, t_{k-1}) \cdot [x_{t_{k-1}} - \bar{X}_{t_{k-1}}^\theta]\right]$$

$$- \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds,$$

$$Q_k(\theta) := \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{q}^\theta(\bar{X}_s^\theta) \cdot (Z^\theta)^T(t_k, s) ds.$$

By Lemma 11 in the appendix we have that the inverse matrices Q_k^{-1} exist.

The rest of this section studies regularity of the MCE when n is held fixed while $\epsilon + \delta \to 0$. Theorem 3 establishes that the MCE is a consistent estimator of the true value θ_0, and Theorem 4 that it is asymptotically normal, with a limiting variance $M(\theta; n)$ which we calculate explicitly. Finally, Lemma 1 shows that this limiting variance attains, in the limit as $n \to \infty$, the Cramér–Rao bound for the continuous-data estimation problem, which is to say that the estimator is asymptotically statistically efficient as first $\epsilon + \delta \to 0$ and then $n \to \infty$.

Before developing the theory, we need to define the limiting contrast function

$$\tilde{U}(\theta; \{x_{t_k}\}_{k=1}^n) := \Sigma_{k=1}^n \left[\tilde{F}_k^T(\theta; \{x_{t_k}\}_{k=1}^n) \cdot Q_k^{-1}(\theta) \cdot \tilde{F}_k(\theta; \{x_{t_k}\}_{k=1}^n) \right],$$

where

$$(15) \qquad \tilde{F}_k(\theta; \{x_{t_k}\}_{k=1}^n) := \left[x_{t_k} - \bar{X}_{t_k}^\theta \right] - Z^\theta(t_k, t_{k-1}) \cdot \left[x_{t_{k-1}} - \bar{X}_{t_{k-1}}^\theta \right].$$

By Lemma 12 in the appendix, we have that

$$\lim_{\epsilon \to 0} P \left(\sup_{\theta_1, \theta_2 \in \Theta} |U^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| > \eta \right) = 0.$$

Denoting by θ_0 the true value of the unknown parameter, we must assume an appropriate identifiability condition to guarantee that $\tilde{U}(\theta; \{\bar{X}_{t_k}^{\theta_0}\}_{k=1}^n)$ is uniquely minimized at $\theta = \theta_0$.

CONDITION 5 (first identifiability condition). *For any $\theta_0 \in \Theta$, $\bar{X}_{t_k}^\theta = \bar{X}_{t_k}^{\theta_0}$ for $k = 1, 2, \ldots, n$ if and only if $\theta = \theta_0$.*

Remark 3. Condition 5 implies that $\tilde{F}_k(\theta; \{\bar{X}^{\theta_0}\}_{k=1}^n) = 0$ for $k = 1, 2, \ldots, n$ if and only if $\theta = \theta_0$ and, hence, that $\tilde{U}(\theta; \{\bar{X}_{t_k}^{\theta_0}\}_{k=1}^n)$ is uniquely minimized at $\theta = \theta_0$.

THEOREM 3 (consistency of the MCE). *Assume Conditions* 1, 2, 4, *and* 5, *and, in the ∞ regime, Condition* 3. *For any $\theta_0 \in \Theta$ and $\eta > 0$,*

$$\lim_{\epsilon \to 0} P \left(|\bar{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon, \theta_0}\}_{k=1}^n) - \theta_0| > \eta \right) = 0.$$

Proof. Consider the modulus of continuity

$$w^\varepsilon(\{x_{t_k}\}_{k=1}^n, \phi) := \sup_{\theta_1, \theta_2 \in \Theta; |\theta_1 - \theta_2| \leq \phi} |U^\varepsilon(\theta_1; \{x_{t_k}\}_{k=1}^n) - U^\varepsilon(\theta_2; \{x_{t_k}\}_{k=1}^n)|.$$

By Theorem 3.2.8 in [6], it suffices to show that $w^\varepsilon(\{X_{t_k}^{\varepsilon, \theta}\}_{k=1}^n, \phi)$ converges in probability uniformly in ϕ to a function of ϕ that tends to 0 as $\phi \to 0$. In fact,

$$\lim_{\epsilon \to 0} P \left(\sup_{\theta \in \Theta} \sup_{\phi > 0} |w^\varepsilon(\{X_{t_k}^{\varepsilon, \theta}\}_{k=1}^n, \phi) - \tilde{w}^\theta(\phi)| > \eta \right) = 0,$$

where

$$\tilde{w}^\theta(\phi) := \sup_{\theta_1, \theta_2 \in \Theta; |\theta_1 - \theta_2| \leq \phi} \left| \tilde{U}(\theta_1; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) \right|;$$

this is immediate by Lemma 12 upon writing

$$\left(U^\varepsilon(\theta_1; \{X_{t_k}^{\varepsilon, \theta}\}_{k=1}^n) - U^\varepsilon(\theta_2; \{X_{t_k}^{\varepsilon, \theta}\}_{k=1}^n) \right) - \left(\tilde{U}(\theta_1; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) \right)$$

$$= \left(U^\varepsilon(\theta_1; \{X_{t_k}^{\varepsilon, \theta}\}_{k=1}^n) - \tilde{U}(\theta_1; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) \right) - \left(U^\varepsilon(\theta_2; \{X_{t_k}^{\varepsilon, \theta}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) \right).$$

Thus $\lim_{\phi \to 0} \tilde{w}^\theta(\phi) = 0$ is a corollary of Lemma 13 in the appendix, concluding the proof of the theorem. $\qquad\square$

Before proceeding to a central limit theorem for the MCE, we must assume an appropriate identifiability condition to guarantee that the variance is well behaved in the limit.

CONDITION 6 (second identifiability condition). *For any $\theta \in \Theta$, the weighted difference*

$$Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}^\theta_{t_{k-1}} - \nabla_\theta \bar{X}^\theta_{t_k}$$

is nonzero for at least one value of $k = 1, 2, \ldots, n$.

THEOREM 4 (asymptotic normality of the MCE). *Assume Conditions 1, 2, 4, 5, and 6, and, in the ∞ regime, Condition 3. For any given $\theta_0 \in \Theta$,*

$$\frac{1}{\sqrt{\epsilon}}(\bar{\theta}^\varepsilon(\{X^{\varepsilon,\theta_0}_{t_k}\}^n_{k=1}) - \theta_0)$$

converges in distribution as $\epsilon \to 0$ to the normal distribution $\mathcal{N}(0, M(\theta_0))$, where the covariance is given by the formula

$$M(\theta) := \left[\Sigma^n_{k=1}\left[\left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}^\theta_{t_{k-1}} - \nabla_\theta \bar{X}^\theta_{t_k}\right)^T\right.\right.$$
$$\left.\left. \cdot Q^{-1}_k(\theta) \cdot \left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}^\theta_{t_{k-1}} - \nabla_\theta \bar{X}^\theta_{t_k}\right)\right]\right]^{-1}.$$

Proof. Let us suppress the data $\{X^{\varepsilon,\theta_0}_{t_k}\}^n_{k=1}$. By Taylor's theorem,

$$0 = \frac{1}{\sqrt{\epsilon}}(\nabla_\theta U^\varepsilon)(\bar{\theta}^\varepsilon) = \frac{1}{\sqrt{\epsilon}}(\nabla_\theta U^\varepsilon)(\theta_0) + \frac{1}{\sqrt{\epsilon}}(\nabla^2_\theta U^\varepsilon)(\theta^{\varepsilon,\dagger}) \cdot (\bar{\theta}^\varepsilon - \theta_0),$$

where $\theta^{\varepsilon,\dagger}$ is an appropriately chosen point on the segment connecting $\bar{\theta}^\varepsilon$ with θ_0. Assuming the inverse exists, we may reexpress this as

$$(16) \qquad \frac{1}{\sqrt{\epsilon}}(\bar{\theta}^\varepsilon - \theta_0) = (\nabla^2_\theta U^\varepsilon)^{-1}(\theta^{\varepsilon,\dagger}) \cdot \frac{1}{\sqrt{\epsilon}}(\nabla_\theta U^\varepsilon)(\theta_0).$$

Thus, it suffices to establish a limit in distribution of $\frac{1}{\sqrt{\epsilon}}(\nabla_\theta U^\varepsilon)(\theta_0)$ and an invertible limit in probability of $(\nabla^2_\theta U^\varepsilon)(\theta^{\varepsilon,\dagger})$; the interested reader is referred to section 3.3.4 in [6] for a rigorous justification of this now-classical approach.

For the weighted gradient of the contrast function, we have that

$$(17) \qquad \frac{1}{\sqrt{\epsilon}}(\nabla_\theta U^\varepsilon)(\theta_0) = [\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}]\Big|_{\theta=\theta_0},$$

where

$$\mathrm{I} := \frac{2}{\sqrt{\epsilon}}\Sigma^n_{k=1}\left[\left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}^\theta_{t_{k-1}} - \nabla_\theta \bar{X}^\theta_{t_k}\right)^T \cdot Q^{-1}_k(\theta) \cdot F^\varepsilon_k(\theta)\right],$$

$$\mathrm{II} := -\frac{2}{\sqrt{\epsilon}}\Sigma^n_{k=1}\left[\left(\nabla_\theta Z^\theta(t_k, t_{k-1}) \cdot \left[X^{\varepsilon,\theta_0}_{t_{k-1}} - \bar{X}^\theta_{t_{k-1}}\right]\right)^T \cdot Q^{-1}_k(\theta) \cdot F^\varepsilon_k(\theta)\right],$$

$$\mathrm{III} := -2\Sigma^n_{k=1}\left[\left(\nabla_\theta \int^{t_k}_{t_{k-1}} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}^\theta_s)ds\right)^T \cdot Q^{-1}_k(\theta) \cdot F^\varepsilon_k(\theta)\right],$$

$$\mathrm{IV} := \frac{1}{\sqrt{\epsilon}}\Sigma^n_{k=1}\left[(F^\varepsilon_k)^T(\theta) \cdot \nabla_\theta\left(Q^{-1}_k(\theta)\right) \cdot F^\varepsilon_k(\theta)\right].$$

Note that $||\nabla_\theta Z^\theta(t_k, t_{k-1})||$ is uniformly bounded, that $\frac{1}{\sqrt{\epsilon}} E|X_{t_{k-1}}^{\varepsilon,\theta_0} - \bar{X}_{t_{k-1}}^{\theta_0}|$ is bounded for ϵ sufficiently small by Theorem 1, that $||Q_k^{-1}(\theta)||$ is bounded by Lemma 11, and that $E|F_k^\varepsilon(\theta)|$ vanishes at the rate of $\sqrt{\epsilon}$ by Lemma 8; putting these together, we deduce that $\mathrm{II}\big|_{\theta=\theta_0}$ vanishes in probability. $\mathrm{III}\big|_{\theta=\theta_0}$ and $\mathrm{IV}\big|_{\theta=\theta_0}$ similarly vanish in probability. Meanwhile, $\mathrm{I}\big|_{\theta=\theta_0}$ converges in distribution to $\mathcal{N}(0, \Gamma(\theta_0))$ with

$$(18) \qquad \Gamma(\theta) = 4\Sigma_{k=1}^n \bigg[\Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big)^T$$
$$\cdot Q_k^{-1}(\theta) \cdot \Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big) \bigg].$$

To see this, let ψ stand for χ in the ∞ regime and for Φ in the γ regime. Recalling the definition (14) of F_k^ε we write, with $\tilde{\mathcal{R}}^\varepsilon$ as in Theorem 2,

$$(19)$$
$$F_k^\varepsilon(\theta) = \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s)(\sigma + \nabla_y \psi \cdot \tau_1)(X_s^{\varepsilon,\theta}, Y_s^{\epsilon,\theta}) dW_s$$
$$+ \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \nabla_y \psi \cdot \tau_2 (X_s^{\varepsilon,\theta}, Y_s^{\epsilon,\theta}) dB_s + \sqrt{\epsilon} \Big[\tilde{\mathcal{R}}_{t_k}^{\varepsilon,\theta} - Z^\theta(t_k, t_{k-1}) \tilde{\mathcal{R}}_{t_{k-1}}^{\varepsilon,\theta} \Big].$$

Thus, we see that the sequence $\{\frac{1}{\sqrt{\epsilon}} F_k^\varepsilon(\theta)\}_{k=1}^n$ converges in distribution, as a sequence, to the sequence $\{\int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \bar{q}^{1/2}(\bar{X}_s^\theta) d(W, B)_s\}_{k=1}^n$ of independent Gaussian random variables, whence I converges in distribution to

$$2\Sigma_{k=1}^n \int_{t_{k-1}}^{t_k} \Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big)^T \cdot Q_k^{-1}(\theta) \cdot Z^\theta(t_k, s) \bar{q}^{1/2}(\bar{X}_s^\theta) d(W, B)_s,$$

which is of course a centered Gaussian random variable with covariance matrix

$$4\Sigma_{k=1}^n \int_{t_{k-1}}^{t_k} \bigg[\Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big)^T \cdot Q_k^{-1}(\theta) \cdot Z^\theta(t_k, s) \bar{q}^{1/2}(\bar{X}_s^\theta) \bigg]$$
$$\times \bigg[\Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big)^T \cdot Q_k^{-1}(\theta) \cdot Z^\theta(t_k, s) \bar{q}^{1/2}(\bar{X}_s^\theta) \bigg]^T ds$$
$$= 4\Sigma_{k=1}^n \bigg[\Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big)^T \cdot Q_k^{-1}(\theta) \bigg] \cdot Q_k(\theta)$$
$$\times \bigg[\Big(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \Big)^T \cdot Q_k^{-1}(\theta) \bigg]^T,$$

which is exactly $\Gamma(\theta)$. For the Hessian of the contrast function, we have, in the notation of (17),

$$(20) \qquad (\nabla_\theta^2 U^\varepsilon)(\theta^{\varepsilon,\dagger}) = \sqrt{\epsilon} \Big[\nabla_\theta (\mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV}) \Big] \Big|_{\theta=\theta_0}.$$

The terms $\sqrt{\epsilon}\nabla_\theta\mathrm{II}\big|_{\theta=\theta_0}$, $\sqrt{\epsilon}\nabla_\theta\,\mathrm{III}\big|_{\theta=\theta_0}$, and $\sqrt{\epsilon}\nabla_\theta\,\mathrm{IV}\big|_{\theta=\theta_0}$ vanish in probability. Meanwhile, the first term may be rewritten as

$$
\begin{aligned}
\sqrt{\epsilon}\nabla_\theta I\big|_{\theta=\theta_0} = {} & 2\Sigma_{k=1}^n\left[\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)^T\cdot Q_k^{-1}(\theta)\right.\\
& \left.\left.\cdot\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)\right]\right|_{\theta=\theta_0}\\
& - 2\Sigma_{k=1}^n\left[\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)^T\cdot Q_k^{-1}(\theta)\right.\\
& \left.\left.\cdot\left(\nabla_\theta Z^\theta(t_k,t_{k-1})\cdot\left[X_{t_{k-1}}^{\varepsilon,\theta_0} - \bar{X}_{t_{k-1}}^\theta\right]\right)\right]\right|_{\theta=\theta_0}\\
& - \sqrt{\epsilon}\cdot 2\Sigma_{k=1}^n\left[\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)^T\right.\\
& \left.\left.\cdot Q_k^{-1}(\theta)\cdot\nabla_\theta\int_{t_{k-1}}^{t_k} Z^\theta(t_k,s)\cdot\bar{J}^\theta(\bar{X}_s^\theta)ds\right]\right|_{\theta=\theta_0}\\
& + \Sigma_{k=1}^n\left[\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)^T\cdot\nabla_\theta\left(Q_k^{-1}(\theta)\right)\cdot F_k^\varepsilon(\theta)\right]\Big|_{\theta=\theta_0}\\
& + 2\Sigma_{k=1}^n\left[\nabla_\theta\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)^T\cdot Q_k^{-1}(\theta)\cdot F_k^\varepsilon(\theta)\right]\Big|_{\theta=\theta_0}.
\end{aligned}
$$

All but the first summand vanish in probability by Theorem 1 and Lemmas 8 and 11 in the appendix; the limit in probability is therefore

$$
\begin{aligned}
2\Sigma_{k=1}^n\left[\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)\right.\\
\left.\left.\cdot Q_k^{-1}(\theta)\cdot\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)\right]\right|_{\theta=\theta_0},
\end{aligned}
$$

which is invertible by Lemma 11 and Condition 6.

Recalling (16) and (19), we conclude that $\frac{1}{\sqrt{\epsilon}}(\bar{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0)$ converges in distribution as $\epsilon\to 0$ to $\mathcal{N}(0,M(\theta_0))$ with

$$
\begin{aligned}
M(\theta) = \left[\Sigma_{k=1}^n\left[\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)^T\right.\right.\\
\left.\left.\cdot Q_k^{-1}(\theta)\cdot\left(Z^\theta(t_k,t_{k-1})\cdot\nabla_\theta\bar{X}_{t_{k-1}}^\theta - \nabla_\theta\bar{X}_{t_k}^\theta\right)\right]\right]^{-1},
\end{aligned}
$$

concluding the proof of the theorem. □

We conclude this section by computing the limit $\lim_{n\to\infty} M(\theta;n)$, where the notation $M(\theta;n) = M(\theta)$ makes explicit the dependence on the number of data points n. A calculation shows that, in light of the ergodic theorems in the appendix, the Fisher information established in Theorem 3.3.1 in [17] leads to the asymptotic

Fisher information

$$\int_0^T \left(\nabla_\theta \bar\lambda^\theta\right)^T (\bar X_s^\theta) \cdot (\bar q^\theta)^{-1}(\bar X_s^\theta) \cdot \left(\nabla_\theta \bar\lambda^\theta\right)(\bar X_s^\theta)ds$$

for the continuous-data estimation problem. Thus, Lemma 1 shows in particular that $M(\theta;n)$ attains, in the limit as $n \to \infty$, the Cramér–Rao bound for the continuous-data estimation problem, which is to say that the estimator is asymptotically statistically efficient as first $\epsilon + \delta \to 0$ and then $n \to \infty$.

LEMMA 1. *Assume Conditions* 1, 2, 4, 5, *and* 6 *and, in the* ∞ *regime, Condition* 3. *For any given* $\theta \in \Theta$,

$$\lim_{n\to\infty} M(\theta;n) = \left[\int_0^T \left(\nabla_\theta \bar\lambda^\theta\right)^T (\bar X_s^\theta) \cdot (\bar q^\theta)^{-1}(\bar X_s^\theta) \cdot \left(\nabla_\theta \bar\lambda^\theta\right)(\bar X_s^\theta)ds\right]^{-1},$$

where $M(\theta;n)$ *is the limiting variance with* n *samples as per Theorem* 4.

Proof. It is equivalent to show convergence of the inverses; that is, to show that

$$\lim_{n\to\infty} M^{-1}(\theta;n) = \int_0^T \left(\nabla_\theta \bar\lambda^\theta\right)^T (\bar X_s^\theta) \cdot (\bar q^\theta)^{-1}(\bar X_s^\theta) \cdot \left(\nabla_\theta \bar\lambda^\theta\right)(\bar X_s^\theta)ds.$$

Notice that

$$(21) \qquad Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar X_{t_{k-1}}^\theta - \nabla_\theta \bar X_{t_k}^\theta$$

$$= \left(Z^\theta(t_k, t_{k-1}) - 1 - \int_{t_{k-1}}^{t_k} (\nabla_x \bar\lambda^\theta)(\bar X_u^\theta)du\right) \cdot \nabla_\theta \bar X_{t_{k-1}}^\theta$$

$$+ \int_{t_{k-1}}^{t_k} (\nabla_x \bar\lambda^\theta)(\bar X_u^\theta)du \cdot \nabla_\theta \bar X_{t_{k-1}}^\theta - \nabla_\theta \left(\bar X_{t_k}^\theta - \bar X_{t_{k-1}}^\theta\right)$$

$$= \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where, with 1_m being the $m \times m$ identity matrix,

$$\mathrm{I} := \left(Z^\theta(t_k, t_{k-1}) - 1_m - \int_{t_{k-1}}^{t_k} (\nabla_x \bar\lambda^\theta)(\bar X_u^\theta)du\right) \cdot \nabla_\theta \bar X_{t_{k-1}}^\theta,$$

$$\mathrm{II} := \int_{t_{k-1}}^{t_k} (\nabla_x \bar\lambda^\theta)(\bar X_u^\theta) \cdot \nabla_\theta \left(\bar X_{t_{k-1}}^\theta - \bar X_u^\theta\right) du,$$

$$\mathrm{III} := -\int_{t_{k-1}}^{t_k} (\nabla_\theta \bar\lambda^\theta)(\bar X_u^\theta)du.$$

By Taylor's theorem, there is a constant K such that $|I| \le K \cdot \Delta^2$; likewise, there is a constant K such that $|\mathrm{II}| \le K \cdot \Delta^2$. Meanwhile, the exponential terms in the definition of Q converge uniformly to the identity. Finally, $\mathrm{III} := -\int_{t_{k-1}}^{t_k} (\nabla_\theta \bar\lambda^\theta)(\bar X_u^\theta)du$ behaves like $-\Delta \cdot (\nabla_\theta \bar\lambda^\theta)(\bar X_{t_{k-1}}^\theta)$, whence we conclude that $M^{-1}(\theta;n)$ behaves asymptotically as the Riemann sum

$$\Delta \cdot \Sigma_{k=1}^n \left[(\nabla_\theta \bar\lambda^\theta)^T \cdot (\bar q^\theta)^{-1} \cdot (\nabla_\theta \bar\lambda^\theta)\right](\bar X_{t_{k-1}}^\theta),$$

whose limit is of course $\int_0^T [(\nabla_\theta \bar\lambda^\theta)^T \cdot (\bar q^\theta)^{-1} \cdot (\nabla_\theta \bar\lambda^\theta)](\bar X_s^\theta)ds$, concluding the proof of the lemma. ▯

5. A simplified estimator. In this section we show that consistency and asymptotic normality are still achieved, even if one omits the covariance weights Q_k^{-1} in the

contrast function. On the one hand, one incurs by this omission an increase in the
limiting variance, and asymptotic efficiency is lost (see Lemma 3). On the other
hand, the simplification affords certain advantages—for instance, one may compute
the *simplified* contrast even if σ is unknown. The simplified estimation procedure also
presents enhanced robustness in numerical simulations (see section 7), showing less
sensitivity to the values of ϵ and δ, which are commonly unknown in practice.

With the omission of the weights, one essentially imposes the additional sim-
plification of constant diffusion on the misspecified model that gave rise to the first
contrast estimator (12). Accordingly, the proofs of the results presented in this section
follow near verbatim the arguments of section 4, and will not be presented separately
in this section. The identifiability conditions remain unchanged. We also substitute
from the outset the $\tilde{F}_k$ of (15) for the F_k^ε of (14); it is not hard to see that the integral
terms lost in doing so are in any event asymptotically negligible.

Let us therefore define the SMCE

$$(22) \qquad \tilde{\theta}(\{x_{t_k}\}_{k=1}^n) := \mathrm{argmin}_{\theta \in \Theta} \Sigma_{k=1}^n |\tilde{F}_k(\theta; \{x_{t_k}\}_{k=1}^n)|^2.$$

THEOREM 5 (consistency of the SMCE). *Assume Conditions* 1, 2, 4, 5, *and* 6
and, in the ∞ *regime, Condition* 3. *For any* $\theta_0 \in \Theta$ *and* $\eta > 0$,

$$\lim_{\epsilon \to 0} P\left(|\tilde{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0| > \eta\right) = 0.$$

THEOREM 6 (asymptotic normality of the SMCE). *Assume Conditions* 1, 2, 4,
5, *and* 6 *and, in the* ∞ *regime, Condition* 3. *For any given* $\theta_0 \in \Theta$,

$$\frac{1}{\sqrt{\epsilon}}(\tilde{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0)$$

converges in distribution as $\epsilon \to 0$ *to the normal distribution* $\mathcal{N}(0, \tilde{M}(\theta_0))$, *where the
covariance is given by the formula*

$$\tilde{M}(\theta) := \Psi^{-1}(\theta) \cdot \Xi(\theta) \cdot \Psi^{-1}(\theta),$$

where

$$\Psi(\theta) := \Sigma_{k=1}^n \left[\left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \right)^T \right.$$
$$\left. \cdot \left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \right) \right],$$

$$\Xi(\theta) := \Sigma_{k=1}^n \left[\left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \right)^T \right.$$
$$\left. \cdot Q_k(\theta) \cdot \left(Z^\theta(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}}^\theta - \nabla_\theta \bar{X}_{t_k}^\theta \right) \right].$$

LEMMA 2. *Assume Conditions* 1, 2, 4, 5, *and* 6 *and, in the* ∞ *regime, Condition*
3. *For any given* $\theta \in \Theta$,

$$\lim_{n \to \infty} \tilde{M}^{-1}(\theta; n) = \check{\Psi}(\theta) \cdot \check{\Xi}^{-1}(\theta) \cdot \check{\Psi}(\theta),$$

where

$$\check{\Psi}(\theta) := \int_0^T \left[\left(\nabla_\theta \bar{\lambda}^\theta \right)^T \cdot \left(\nabla_\theta \bar{\lambda}^\theta \right) \right] (\bar{X}_s^\theta) ds,$$

$$\check{\Xi}(\theta) := \int_0^T \left[\left(\nabla_\theta \bar{\lambda}^\theta \right)^T \cdot \bar{q}^\theta \cdot \left(\nabla_\theta \bar{\lambda}^\theta \right) \right] (\bar{X}_s^\theta) ds.$$

The next lemma establishes that the variance of the SMCE is bounded below by that of the MCE in the sense that the difference of the covariance matrices is positive semidefinite. In general, one loses asymptotic statistical efficiency when one moves from the MCE to the SMCE.

LEMMA 3. *Assume Conditions* 1, 2, 4, 5, *and* 6 *and, in the* ∞ *regime, Condition* 3. *Let* $M(\theta; n)$ *and* $\tilde{M}(\theta; n)$ *be as in Theorems* 4 *and* 6, *respectively. For any* $\theta \in \Theta$ *and* $n \geq 1$, *the* $k \times k$ *matrix*

$$(23) \qquad \tilde{M}(\theta; n) - M(\theta; n)$$

is positive semidefinite. In particular, for any $\theta \in \Theta$, *the difference of limits*

$$(24) \qquad \lim_{n\to\infty} \tilde{M}(\theta; n) - \lim_{n\to\infty} M(\theta; n)$$

is positive semidefinite.

Proof. We suppress the parameters θ and n and define

$$d_k := Z(t_k, t_{k-1}) \cdot \nabla_\theta \bar{X}_{t_{k-1}} - \nabla_\theta \bar{X}_{t_k}$$

so that, recalling the notation Ψ and Ξ from Theorem 6,

$$\Psi = \Sigma_k d_k^T d_k,$$
$$\Xi = \Sigma_k d_k^T Q_k d_k.$$

The first statement, that $\tilde{M} - M$ is positive semidefinite, is equivalent to the statement that $\tilde{M}^{-1} - M^{-1}$ is negative semidefinite; we now demonstrate the latter. Let $\xi \in \mathbb{R}^k \setminus \{0\}$ be given.

$$
\begin{aligned}
|\xi^T \tilde{M}^{-1}\xi|^2 &= |\Sigma_k \langle d_k \Xi^{-1}\Psi\xi, d_k\xi\rangle|^2 \\
&= |\Sigma_k \langle (Q^{1/2})d_k \Xi^{-1}\Psi\xi, (Q^{1/2})^{-1}d_k\xi\rangle|^2 \\
&\leq \left[\Sigma_k |(Q^{1/2})d_k \Xi^{-1}\Psi\xi|^2\right] \cdot \left[\Sigma_k |(Q^{1/2})^{-1}d_k\xi|^2\right] \\
&= \left[\Sigma_k \langle d_k^T Q d_k \Xi^{-1}\Psi\xi, \Xi^{-1}\Psi\xi\rangle\right] \cdot \left[\Sigma_k \langle d_k^T Q d_k\xi, \xi\rangle\right] \\
&= \left[\xi^T \tilde{M}^{-1}\xi\right] \cdot \left[\xi^T M^{-1}\xi\right].
\end{aligned}
$$

Cancelling a factor of $\xi^T \tilde{M}^{-1}\xi$, which is positive, one obtains

$$\xi^T \left[\tilde{M}^{-1} - M^{-1}\right] \xi \leq 0.$$

This establishes the first statement, that (23) is positive semidefinite. The second statement, that the difference of limits (24) is also positive semidefinite, follows immediately by continuity. □

6. Limit of high-frequency observation. In section 4, we considered the important case in which the number of observations n is fixed, obtaining consistent and asymptotically normal estimators. The question naturally arises as to whether these properties carry over into the asymptotic regime of high-frequency observation, in which n is taken arbitrarily large *at the same time* that ϵ is taken to vanish; that is, recalling our notation $\Delta := T/n$ for the sampling interval, we wish to consider the behavior of the estimators in the joint limit $\epsilon + \Delta \to 0$.

It turns out that if ϵ is $o(\Delta)$ as $\Delta \to 0$, then the theory of sections 4 and 5 carries over with only minor adjustments.

CONDITION 7 (third identifiability condition). *For any $\theta_0 \in \Theta$, the integral*

$$\Lambda_t(\theta, \theta_0) := \bar{\lambda}^{\theta_0}(\bar{X}_t^{\theta_0}) - \bar{\lambda}^{\theta}(\bar{X}_t^{\theta}) - (\nabla_x \bar{\lambda}^{\theta})(\bar{X}_t^{\theta}) \cdot [\bar{X}_t^{\theta_0} - \bar{X}_t^{\theta}]$$

vanishes for all $t \in [0, T]$ if and only if $\theta = \theta_0$.

THEOREM 7 (consistency of the MCE and SMCE as $\epsilon + \Delta \to 0$). *Assume Conditions 1, 2, 4, and 7 and, in the ∞ regime, Condition 3. Assume that Δ does not decrease too quickly relative to ϵ, so that ϵ is $o(\Delta)$ as $\Delta \to 0$. For any $\theta_0 \in \Theta$ and $\eta > 0$,*

$$\lim_{(\epsilon+\Delta) \to 0} P\left(|\bar{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0| > \eta\right) = 0, \quad \lim_{(\epsilon+\Delta) \to 0} P\left(|\tilde{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0| > \eta\right) = 0.$$

Proof. The proof is similar to those of Theorems 3 and 5, with Lemma 16 in place of Lemma 12. ☐

The following identifiability condition is the limiting analogue of Condition 6.

CONDITION 8 (fourth identifiability condition). *For any $\theta \in \Theta$, the matrix-valued integral*

$$(25) \qquad \int_0^T [(\nabla_\theta \bar{\lambda}^\theta)^T (\nabla_\theta \bar{\lambda}^\theta)](\bar{X}_s^\theta) ds$$

is invertible.

We wish to mimic the proof of Theorem 4. When one is considering the joint limit $\epsilon + \Delta \to 0$, the contribution of the covariance corrections Q_k^{-1} to the θ derivatives of the contrast function becomes a rather delicate quantity to handle. We therefore state and prove asymptotic normality of the MCE in this joint limit, only under the stronger assumption that ϵ is $o(\Delta^2)$, rather than merely $o(\Delta)$, as $\Delta \to 0$. This is Theorem 8.

Notably, the stronger assumption is not needed to adapt with little modification the proof of asymptotic normality of the SMCE for the same joint limit. With the omission of the weights, one essentially imposes the additional simplification of constant diffusion on the misspecified model that motivates the estimator—in particular, the corrections Q_k^{-1} are essentially replaced with $1/\Delta$, and the delicate θ dependence is avoided altogether. Note that Theorem 9 assumes only that ϵ is $o(\Delta)$.

THEOREM 8 (asymptotic normality of the MCE as $\epsilon + \Delta \to 0$). *Assume Conditions 1, 2, 4, 7, and 8 and, in the ∞ regime, Condition 3. Assume that Δ decreases slowly relative to ϵ, so that ϵ is $o(\Delta^2)$ as $\Delta \to 0$. For any given $\theta_0 \in \Theta$, $\frac{1}{\sqrt{\epsilon}}(\bar{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0)$ converges in distribution as $\epsilon + \Delta \to 0$ to the normal distribution $\mathcal{N}(0, M(\theta_0))$, where the covariance is given by the formula*

$$M(\theta) := \left[\int_0^T \left(\nabla_\theta \bar{\lambda}^\theta\right)^T (\bar{X}_s^\theta) \cdot (\bar{q}^\theta)^{-1}(\bar{X}_s^\theta) \cdot \left(\nabla_\theta \bar{\lambda}^\theta\right)(\bar{X}_s^\theta) ds \right]^{-1}.$$

Proof. Let us examine the proof of Theorem 4. Consider, in the notation of (17), the term II. It is easy to see that $\nabla_\theta Z^\theta(t_k, t_{k-1})$ is uniformly bounded by a constant times Δ, and Lemma 11 establishes that $||Q_k^{-1}(\theta)||$ is bounded by a constant times $1/\Delta$. Meanwhile, $E|\frac{1}{\sqrt{\epsilon}}[X_{t_{k-1}}^{\varepsilon,\theta_0} - \bar{X}_{t_{k-1}}^{\theta}]|^2$ is uniformly bounded by Theorem 1. Thus,

by an application of Hölder's inequality, II will vanish even as $n \to \infty$, provided that the general term $E|F_k^\varepsilon(\theta)|^2$ vanishes faster than Δ^2; but this is immediate from Lemma 14, since we have assumed that ϵ tends to 0 faster than does Δ. III vanishes similarly. IV vanishes by Lemma 14 and the fact that $||\nabla_\theta(Q_k^{-1}(\theta))||$ is bounded by a constant times $1/\Delta$ (it is in order to accommodate this somewhat naïve bound on the derivative of the covariance correction that we have assumed for this theorem that ϵ is $o(\Delta^2)$, rather than merely $o(\Delta)$). The convergence in distribution of I is as before, although in the regime of joint asymptotics $\epsilon + \Delta \to 0$, one must of course combine the arguments of Theorem 4 and Lemma 1.

The convergence in probability of the Hessian (20) even as $n \to \infty$ is likewise. □

THEOREM 9 (asymptotic normality of the SMCE as $\epsilon + \Delta \to 0$). *Assume Conditions* 1, 2, 4, 7, *and* 8 *and, in the* ∞ *regime, Condition* 3. *Assume that* Δ *does not decrease too quickly relative to* ϵ, *so that* ϵ *is* $o(\Delta)$ *as* $\Delta \to 0$. *For any given* $\theta_0 \in \Theta$, $\frac{1}{\sqrt{\epsilon}}(\tilde{\theta}^\varepsilon(\{X_{t_k}^{\varepsilon,\theta_0}\}_{k=1}^n) - \theta_0)$ *converges in distribution as* $\epsilon + \Delta \to 0$ *to the normal distribution* $\mathcal{N}(0, \tilde{M}(\theta_0))$, *where the covariance is given by the formula*

$$\tilde{M}(\theta) := \check{\Psi}^{-1}(\theta) \cdot \check{\Xi}(\theta) \cdot \check{\Psi}^{-1}(\theta),$$

where

$$\check{\Psi}(\theta) := \int_0^T \left[\left(\nabla_\theta \bar{\lambda}^\theta\right)^T \cdot \left(\nabla_\theta \bar{\lambda}^\theta\right) \right] (\bar{X}_s^\theta) ds,$$

$$\check{\Xi}(\theta) := \int_0^T \left[\left(\nabla_\theta \bar{\lambda}^\theta\right)^T \cdot \bar{q}^\theta \cdot \left(\nabla_\theta \bar{\lambda}^\theta\right) \right] (\bar{X}_s^\theta) ds.$$

Proof. As before, the arguments for the SMCE follow nearly verbatim those for the MCE. The weaker assumption $\epsilon = o(\Delta)$ suffices in this case because the term for which it was needed in the proof of Theorem 8, i.e., term IV in (17), does not appear if one is using the simplified, in place of the original, contrast function. □

Remark 4. It is clear that one can reexpress the conditions $\epsilon = o(\Delta)$ and $\epsilon = o(\Delta^2)$ in terms of δ and Δ. We have deliberately chosen instead to present the results in terms of ϵ and Δ because Δ is known in general (via T and n) whereas ϵ can in principle be estimated via the magnitude of the quadratic variation of the process X^ε. We emphasize that in contrast to the rest of the subsampling literature, e.g., [2, 20], we do not (in fact, cannot) impose conditions on how $1/\Delta$ grows relative to n, due simply to the fact that we have *fixed* $T = n \cdot \Delta$ to study the regime of small noise.

7. Numerical examples. We now present data from numerical simulations to supplement and illustrate the theory. We begin by considering the system

(26)
$$dX_t^\varepsilon = \frac{\epsilon}{\delta}(\sin(Y_t^\varepsilon) - \cos(Y_t^\varepsilon))dt - \theta_0 X_t^\varepsilon dt + \sqrt{\epsilon}dW_t,$$
$$Y_t^\varepsilon := X^\varepsilon/\delta$$

for $t \in [0, T = 1]$ with $X_0^\varepsilon = Y_0^\varepsilon = 1 \in \mathbb{R}$. This system was studied as an example in [27]. We suppose that we are in the ∞ regime with $\delta = \epsilon^{3/2}$. The results of sections 4, 5, and 6 on consistency and asymptotic normality are illustrated in the numerical data that follow.

The limit $\bar{X}$ of the slow process X^ε in (26) is

$$\bar{X}_t = e^{-t \cdot \theta_0 \cdot \left(\frac{2\pi}{L}\right)^2},$$

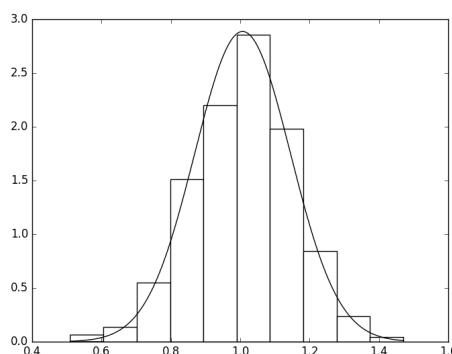where $L := \int_0^{2\pi} e^{2(\sin(y) + \cos(y))} dy$.

TABLE 1

Example 1: MCE of $\theta_0 = 1$ with empirical confidence intervals and theoretical standard deviations.

| ϵ	n	Mean estimator	68% confidence interval	95% confidence interval	Theoretical SD
10^{-2}	10	1.0173	(0.5483, 1.4862)	(0.0794, 1.9552)	0.4370
10^{-2}	10^2	1.0430	(0.5688, 1.5173)	(0.0945, 1.9916)	0.4370
10^{-2}	10^3	1.0686	(0.6104, 1.5267)	(0.1523, 1.9848)	0.4370
10^{-3}	10	1.0005	(0.8610, 1.1401)	(0.7215, 1.2794)	0.1382
10^{-3}	10^2	0.9992	(0.8571, 1.1413)	(0.7150, 1.2834)	0.1382
10^{-3}	10^3	1.0068	(0.8643, 1.1494)	(0.7218, 1.2919)	0.1382

TABLE 2

Example 1: SMCE of $\theta_0 = 1$ with empirical confidence intervals and theoretical standard deviations.

ϵ	n	Mean estimator	68% confidence interval	95% confidence interval	Theoretical SD
10^{-2}	10	1.0710	(0.5902, 1.5518)	(0.1095, 2.0326)	0.4370
10^{-2}	10^2	1.0649	(0.5845, 1.5453)	(0.1041, 2.0256)	0.4370
10^{-2}	10^3	1.0764	(0.5958, 1.5570)	(0.1151, 2.0377)	0.4370
10^{-3}	10	1.0115	(0.8732, 1.1498)	(0.7349, 1.2880)	0.1382
10^{-3}	10^2	1.0074	(0.8598, 1.1549)	(0.7122, 1.3025)	0.1382
10^{-3}	10^3	1.0108	(0.8690, 1.1525)	(0.7272, 1.2943)	0.1382
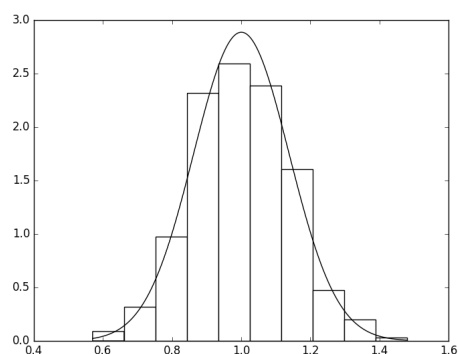
FIG. 1. *Example 1, MCE, $\epsilon = 10^{-3}$, $n = 10$.*   FIG. 2. *Example 1, MCE, $\epsilon = 10^{-3}$, $n = 10^3$.*

We fix $\theta_0 = 1$ and simulate trajectories using an Euler scheme with 10^8 evenly spaced discrete time steps. We simulate both estimators, $\bar{\theta}^\varepsilon$ and $\tilde{\theta}^\varepsilon$, 1000 times for each combination of $\epsilon = 10^{-2}, 10^{-3}$ and $n = 10, 10^2, 10^3$. Tables 1 and 2 present in each case the mean estimate, a normal-based confidence interval using the empirical standard deviation, and the theoretical standard deviation as per Theorems 4 and 6. The histograms in Figures 1–4 compare the empirical distribution of the estimates with the theoretical density curve.

Tables 1 and 2 and Figures 1–4 illustrate that even in the homogenization regime and even with a small number of data points, the MCEs are consistent and asymptotically normal. We point out that although the empirical variance of $\tilde{\theta}^\varepsilon$ was larger than that of $\bar{\theta}^\varepsilon$, the difference was rather small, especially for small values of ϵ—this
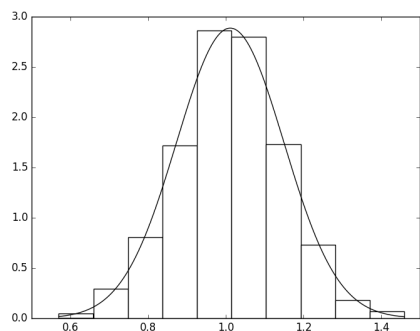
FIG. 3. *Example* 1, *SMCE*, $\epsilon = 10^{-3}$, $n = 10$.
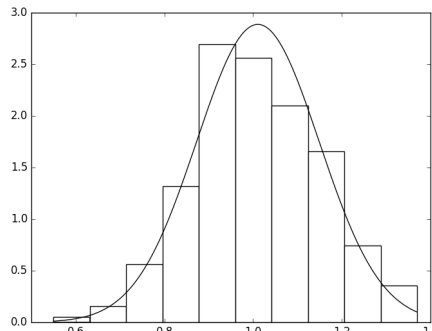
FIG. 4. *Example* 1, *SMCE*, $\epsilon = 10^{-3}$, $n = 10^3$.

is consistent with the fact that the *theoretical* limiting difference in this example was rather negligible. It is also worth pointing out that coefficients depending on θ appear in both slow and fast components.

Notice that the dependence on the fast process in (26) is periodic and that one may therefore interpret the fast process as taking values in a torus. We now consider a second example to illustrate the case where the fast dynamics are not restricted to a compact space:

$$(27) \qquad dX_t^\varepsilon = \frac{\epsilon}{\delta}\theta_0 Y_t^\varepsilon dt + \theta_0 X_t^\varepsilon (Y_t^\varepsilon)^2 dt + \sqrt{\epsilon}dW_t,$$

$$dY_t^\varepsilon = -\frac{\epsilon}{\delta^2}\frac{1}{\theta_0}Y_t^\varepsilon dt + \frac{\sqrt{\epsilon}}{\delta}dB_t$$

for $t \in [0, T = 1]$ with $X_0^\varepsilon = Y_0^\varepsilon = 1 \in \mathbb{R}$. Again, we suppose that we are in the ∞ regime with $\delta = \epsilon^{3/2}$ (this is the more challenging regime, relative to the γ regime). As in the first example, coefficients depending on θ appear in both slow and fast components.

The limit $\bar{X}$ of the slow process X^ε in (27) is

$$\bar{X}_t = e^{t \cdot \frac{\theta_0^2}{2}}.$$

We fix $\theta_0 = 1$ and simulate trajectories using an Euler scheme with 10^8 evenly spaced discrete time steps. We simulate $\bar{\theta}^\varepsilon$ 1000 times for each of $\epsilon = 10^{-2}, 10^{-3}$ with $n = 10$; we simulate $\tilde{\theta}^\varepsilon$ 1000 times for each combination of $\epsilon = 10^{-2}, 10^{-3}$ and $n = 10, 10^2, 10^3$. Tables 3 and 4 present in each case the mean estimate, a normal-based confidence interval using the empirical standard deviation, and the theoretical standard deviation as per Theorem 4. The histograms in Figures 5–8 compare the empirical distribution of the estimates with the theoretical density curve.

Last, we revisit Examples 1 and 2 to see what happens when we deliberately violate the assumption that ϵ is of lower order than $\Delta := T/n$. We repeat some of the simulations above with $n = 10^6$; the results are presented in Tables 5 and 6 and Figures 9–12. One sees that the estimators continue to behave well, suggesting that it may be possible to relax our assumptions about the relationship between ϵ and Δ in the theory of section 6 concerning estimation based on high-frequency observations.

TABLE 3

Example 2: MCE of $\theta_0 = 1$ with empirical confidence intervals and theoretical standard deviations.

ϵ	n	Mean estimator	68% confidence interval	95% confidence interval	Theoretical SD
10^{-2}	10	1.1042	(0.9961, 1.2123)	(0.8879, 1.3205)	0.1079
10^{-3}	10	1.0129	(0.9800, 1.0458)	(0.9472, 1.0787)	0.0341

TABLE 4

Example 2: SMCE of $\theta_0 = 1$ with empirical confidence intervals and theoretical standard deviations.

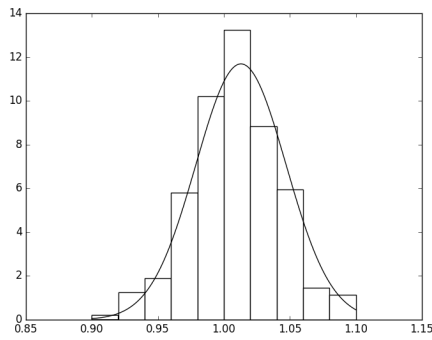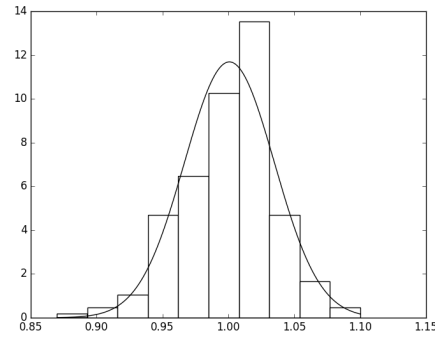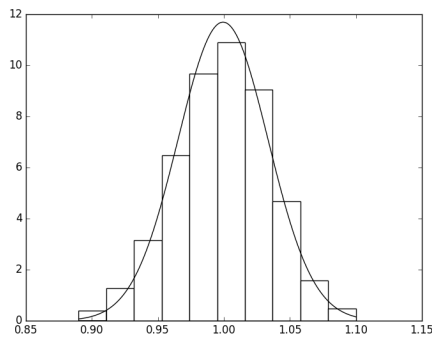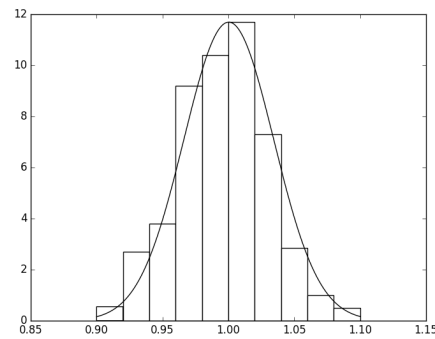| ϵ	n	Mean estimator	68% confidence interval	95% confidence interval	Theoretical SD
10^{-2}	10	0.9840	(0.8704, 1.0977)	(0.7567, 1.2114)	0.1079
10^{-2}	10^2	0.9773	(0.8561, 1.0985)	(0.7349, 1.2197)	0.1079
10^{-2}	10^3	0.9754	(0.8544, 1.0963)	(0.7335, 1.2173)	0.1079
10^{-3}	10	1.0008	(0.9662, 1.0354)	(0.9316, 1.0699)	0.0341
10^{-3}	10^2	0.9994	(0.9654, 1.0333)	(0.9314, 1.0673)	0.0341
10^{-3}	10^3	1.0003	(0.9659, 1.0347)	(0.9314, 1.0692)	0.0341

FIG. 5. *Example 2, MCE, $\epsilon = 10^{-3}$, $n = 10$.*

FIG. 6. *Example 2, SMCE, $\epsilon = 10^{-3}$, $n = 10$.*

FIG. 7. *Example 2, SMCE, $\epsilon = 10^{-3}$, $n = 10^2$.*

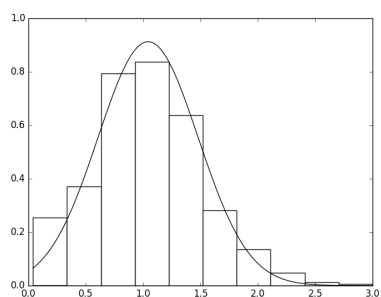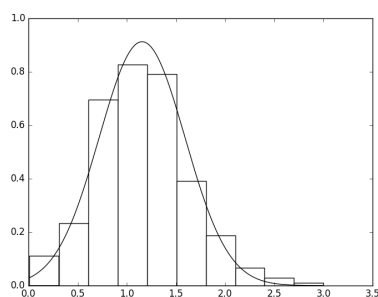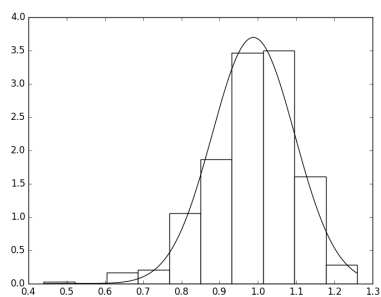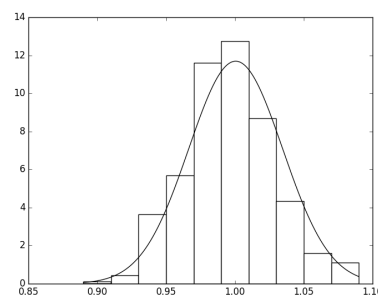FIG. 8. *Example 2, SMCE, $\epsilon = 10^{-3}$, $n = 10^3$.*

TABLE 5
Example 1 revisited: MCE versus SMCE with $\epsilon = 10^{-2}$ and large n ($n = 10^6$).

Estimator	Mean estimator	68% confidence interval	95% confidence interval	Theoretical SD
MCE	1.0427	(0.5630, 1.5225)	(0.0833, 2.0022)	0.437
SMCE	1.1553	(0.6839, 1.6268)	(0.2125, 2.0982)	0.437

TABLE 6
Example 2 revisited: SMCE of $\theta_0 = 1$ with large n ($n = 10^6$).

ϵ	Mean estimator	68% confidence interval	95% confidence interval	Theoretical SD
10^{-2}	0.9883	(0.8777, 1.0988)	(0.7671, 1.2094)	0.1079
10^{-3}	1.0005	(0.9683, 1.0328)	(0.9360, 1.065)	0.0341

FIG. 9. *Example 1, MCE, $\epsilon = 10^{-2}$, $n = 10^6$.*

FIG. 10. *Example 1, SMCE, $\epsilon = 10^{-2}$, $n = 10^6$.*

FIG. 11. *Example 2, SMCE, $\epsilon = 10^{-2}$, $n = 10^6$.*

FIG. 12. *Example 2, SMCE, $\epsilon = 10^{-3}$, $n = 10^6$.*

Remark 5. Data for the MCE $\bar{\theta}^\varepsilon$ for the first example (27) are presented for $n = 10$ only. We have also run simulations with $n = 10^2, 10^3$, etc., but the estimators then began to exhibit a strong positive bias. In contrast, a glance at Table 4 reveals that the SMCE $\tilde{\theta}^\varepsilon$ for the very same model produced robust estimates across all combinations of ϵ and n.

Let us attempt to offer some insight into the reason for this. It turns out that the typical simplified contrast curve in this model increases more steeply as one moves rightward from the minimum $\tilde{\theta}^\varepsilon$ than as one moves leftward. Meanwhile, the effect of introducing the covariance weights Q_k^{-1} in this example is to rescale the contrast

by a factor proportional to $\frac{\theta^2}{1+\theta^4}\frac{1}{e^{\theta\Delta}-1}$. This has the potential to balance the leftward and rightward gradients about the new minimum $\bar{\theta}^\varepsilon$, but at the same time may lead to numerical instabilities when Δ is small. For a given value of ϵ, the convexity of the simplified contrast about its minimum tends to diminish, whence the introduction of the covariance weights may cause the minimum to move too far to the right.

Remark 6. It should be pointed out that the threshold below which ε must fall in order for the theorems to apply will depend upon the problem at hand. The fact that SMCE appears both to be more robust than MCE and at the same time easier to compute (one need not compute the weights Q_k^{-1}) may be viewed as a practical advantage despite the increase in the variance.

8. Conclusions. We have presented in this paper discrete-data estimators for unknown parameters in multiscale diffusion models with small noise. The estimators are defined as minimizers of certain contrast functions motivated by a general second-order stochastic Taylor expansion of the slow process; we have shown them to be consistent, asymptotically normal, and, in the case of the MCE, asymptotically statistically efficient in an appropriate sense.

We also considered the case of high-frequency observation, in which the number of samples n increases to infinity concurrently with the vanishing of the small noise. We showed that the asymptotic properties established for the fixed-n case are still valid provided that the sampling interval $\Delta := T/n$ does not vanish too quickly relative to the small noise. Such conditions are reminiscent of the subsampling prescribed in prior literature on statistical estimation for multiscale models. Despite our best efforts we have not managed to relax them, as has already been done for the case of small noise *without* multiple scales (see [12, 28, 29]). Nevertheless, the numerical simulations of section 7 suggest that improvement is possible, posing an important question for future research to investigate. The crux is the development of an optimal characterization of the rate at which the approximation error $\sqrt{\epsilon}[\tilde{\mathcal{R}}_{t_k}^{\varepsilon,\theta} - Z^\theta(t_k, t_{k-1})\tilde{\mathcal{R}}_{t_{k-1}}^{\varepsilon,\theta}]$ vanishes with respect to $\varepsilon = (\epsilon, \delta)$ and Δ; the δ dependence in particular presents an analytic challenge of unusual delicacy.

Finally, we wish to point out that it seems that the $\mathcal{L}^p$ bounds established in this work can be used to derive Berry–Esseen theorems to characterize the rates of convergence in the central limit theorems.

Appendix. We gather here technical results to which we appeal in the proofs of the main results of this paper.

A.1. Bounds extended from [10]. The SDEs for X^ε and Y^ε that we consider in this paper extend those in [10]. Certain fundamental auxiliary bounds established in [10] are also valid in our case. We gather the necessary bounds in this section. Because the proofs are similar, we do not present them here; the interested reader is referred to [10] for details.

LEMMA 4. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *For any* $p > 0$ *there is a constant* $\tilde{K}$ *such that uniformly in* ϵ *(and hence also* $\delta = \delta(\epsilon)$*) sufficiently small,*

$$E \sup_{0 \le t \le T} |X_t^\varepsilon|^p \le \tilde{K},$$

$$E \sup_{0 \le t \le T} |X_t^\varepsilon - \bar{X}_t|^p \le \tilde{K}.$$

This is similar to Lemma 6 in [10].

Sketch of proof. The only significant difference relative to Lemma 6 in [10] is the introduction of the asymptotically singular term $\frac{\epsilon}{\delta} b(X_t^\varepsilon, Y_t^\varepsilon) dt$ in the ∞ regime. Letting χ be as in (3), applying the Itô formula to $\chi(X_t^\varepsilon, Y_t^\varepsilon)$, and rearranging terms, we obtain

(28)

$$
\int_0^t \frac{\epsilon}{\delta} b(X_s^\varepsilon, Y_s^\varepsilon) ds
$$

$$
= \int_0^t \nabla_y \chi \cdot g(X_s^\varepsilon, Y_s^\varepsilon) ds + \sqrt{\epsilon} \int_0^t (\nabla_y \chi \cdot \tau_1, \nabla_y \chi \cdot \tau_2)(X_s^\varepsilon, Y_s^\varepsilon) d(W, B)_s
$$

$$
+ \epsilon \int_0^t \nabla_x \chi \cdot b(X_s^\varepsilon, Y_s^\varepsilon) ds + \delta \int_0^t \nabla_x \chi \cdot c(X_s^\varepsilon, Y_s^\varepsilon) ds + \epsilon \delta \int_0^t \sigma \sigma^T : \nabla_x^2 \chi(X_s^\varepsilon, Y_s^\varepsilon) ds
$$

$$
+ \epsilon \int_0^t \sigma \tau_1^T : \nabla_y \nabla_x \chi(X_s^\varepsilon, Y_s^\varepsilon) ds + \sqrt{\epsilon} \delta \int_0^t \nabla_x \chi \cdot \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s
$$

$$
- \delta \left(\chi(X_t^\varepsilon, Y_t^\varepsilon) - \chi(x_0, y_0) \right)
$$

$$
=: \int_0^t \nabla_y \chi \cdot g(X_s^\varepsilon, Y_s^\varepsilon) ds + \mathcal{R}_t^\varepsilon;
$$

hence we can write

$$
X_t^\varepsilon = x_0 + \int_0^t \lambda_\infty(X_s^\varepsilon, Y_s^\varepsilon) ds + \sqrt{\epsilon} \int_0^t \sigma(X_s^\varepsilon, Y_s^\varepsilon) dW_s + \mathcal{R}_t^\varepsilon.
$$

The function χ and its various derivatives are bounded uniformly in the first variable by polynomials in the second variable (see Chapters 2 and 3 in [11]). Noting the vanishing prefactors, we conclude that $\mathcal{R}_t^\varepsilon$ vanishes. Thus, the proof may proceed as that of Lemma 6 in [10], with λ_∞ taking the place of c; we omit the details. □

LEMMA 5. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *Let h be a function of x and y satisfying $|h(x,y)| \le K(1 + |x|^r)(1 + |y|^q)$ for some fixed positive constants K, q, r, and let V be either of the Wiener processes W, B. For any $0 < p < \infty$, there is a constant $\tilde{K}$ such that for ϵ (and hence also $\delta = \delta(\epsilon)$) sufficiently small,*

$$
E \int_0^T |h(X_t^\varepsilon, Y_t^\varepsilon)|^p dt \le \tilde{K},
$$

$$
E \sup_{0 \le t \le T} \left| \int_0^t h(X_s^\varepsilon, Y_s^\varepsilon) dV_s \right|^p \le \tilde{K}.
$$

Given Lemma 4, the proof of Lemma 5 is similar to that of Lemma 7 in [10]; we omit the details.

LEMMA 6. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *Let $h(x,y)$ be a function such that $|h(x,y)| \le K(1 + |x|^r)(1 + |y|^q)$ for some fixed positive constants K, q, r, and which, along with each of its derivatives up to second order, is Hölder continuous in y uniformly in x with absolute value growing at most polynomially in $|y|$ as $y \to \infty$. For any $0 < p < \infty$, there is a constant $\tilde{K}$ such that for ϵ (and hence also $\delta = \delta(\epsilon)$) sufficiently small,*

$$
E \sup_{0 \le t \le T} \left| \int_0^t \left(h(X_s^\varepsilon, Y_s^\varepsilon) - \bar{h}(X_s^\varepsilon) \right) ds \right|^p \le \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p,
$$

where $\bar{h}(x)$ is the averaged function $\int_{\mathcal{Y}} h(x,y) \mu_x(dy)$.

Given Lemma 4, the proof of Lemma 6 is similar to that of Theorem 4 in [10]; we omit the details.

LEMMA 7. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *Let* $h(x,y)$ *be a function such that* $|h(x,y)| \leq K(1+|x|^r)(1+|y|^q)$ *for some fixed positive constants* K, q, r, *and which, along with each of its derivatives up to second order, is Hölder continuous in* y *uniformly in* x *with absolute value growing at most polynomially in* $|y|$ *as* $y \to \infty$. *For any* $0 < p < \infty$, *there is a constant* $\tilde{K}$ *such that for* ϵ *(and hence also* $\delta = \delta(\epsilon)$*) sufficiently small,*

$$E \sup_{0 \leq t \leq T} \left| \int_0^t \left(h(X_s^\varepsilon, Y_s^\varepsilon) - \bar{h}(\bar{X}_s) \right) ds \right|^p \leq \tilde{K}(\sqrt{\epsilon} + \sqrt{\delta})^p,$$

where $\bar{h}(x)$ *is the averaged function* $\int_{\mathcal{Y}} h(x,y)\mu_x(dy)$.

Given Theorem 1 and Lemma 6, the proof of Lemma 7 is similar to that of Lemma 10 in [10]; we omit the details.

A.2. Lemmas on F and Q.

LEMMA 8. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *There is a constant* $\tilde{K}$ *such that for* ϵ *(and hence also* $\delta = \delta(\epsilon)$*) sufficiently small,*

$$(29) \qquad E|F_k^\varepsilon(\theta; \{X_{t_k}^{\epsilon,\theta}\}_{k=1}^n)|^p \leq \tilde{K} \cdot \epsilon^{p/2},$$

$$(30) \qquad E|\tilde{F}_k(\theta; \{X_{t_k}^{\epsilon,\theta}\}_{k=1}^n)|^p \leq \tilde{K} \cdot \epsilon^{p/2}.$$

Proof. Recalling that

$$(31) \qquad F_k^\varepsilon(\theta; \{X_{t_k}^{\epsilon,\theta}\}_{k=1}^n) = \left[[X_{t_k}^{\epsilon,\theta} - \bar{X}_{t_k}^\theta] - Z^\theta(t_k, t_{k-1}) \cdot [X_{t_{k-1}}^{\epsilon,\theta} - \bar{X}_{t_{k-1}}^\theta] \right]$$
$$- \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds,$$

the first statement, (29), follows by the triangle inequality and Theorem 1.

(30) may be obtained in the same way by omitting the last term in (32). ☐

In addition to Lemma 8, we need a lemma to establish that $E|F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)|$ and $E|\tilde{F}_k(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)|$ are uniformly bounded in ϵ sufficiently small *even as* θ_1 *and* θ_2 *vary independently.* Of course, with two distinct values of θ, one should not expect the functions to vanish in the limit.

LEMMA 9. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *There is a constant* $\tilde{K}$ *such that for* ϵ *(and hence also* $\delta = \delta(\epsilon)$*) sufficiently small,*

$$(32) \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} E|F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)| \leq \tilde{K},$$

$$(33) \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} E|\tilde{F}_k(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)| \leq \tilde{K}.$$

Proof. Notice that

$$(34) \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} |F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)| \leq \mathrm{I} + \mathrm{II} + \mathrm{III} + \mathrm{IV} + \mathrm{V},$$

where

$$\text{(35)} \qquad \text{I} := \sup_{\theta \in \bar{\Theta}} |X_{t_k}^{\varepsilon,\theta} - X_{t_{k-1}}^{\varepsilon,\theta}|,$$

$$\text{II} := \sup_{\theta \in \bar{\Theta}} |\bar{X}_{t_k}^{\theta} - \bar{X}_{t_{k-1}}^{\theta}| \leq \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |(\nabla_x \bar{\lambda}^{\theta})(\bar{X}_t^{\theta})| \right] \cdot \Delta,$$

$$\text{III} := \sup_{\theta \in \bar{\Theta}} \left| 1_m - Z^{\theta}(t_k, t_{k-1}) \right| \cdot \sup_{\theta \in \bar{\Theta}} |X_{t_{k-1}}^{\varepsilon,\theta}|,$$
$$\leq \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |Z^{\theta}(t, 0) \cdot (\nabla_x \bar{\lambda}^{\theta})(\bar{X}_t^{\theta})| \right] \cdot \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |X_t^{\varepsilon,\theta}| \right] \cdot \Delta,$$

$$\text{IV} := \sup_{\theta \in \bar{\Theta}} \left| 1_m - Z^{\theta}(t_k, t_{k-1}) \right| \cdot \sup_{\theta \in \bar{\Theta}} |\bar{X}_{t_{k-1}}^{\theta}|$$
$$\leq \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |Z^{\theta}(t, 0) \cdot (\nabla_x \bar{\lambda}^{\theta})(\bar{X}_t^{\theta})| \right] \cdot \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |\bar{X}_t^{\theta}| \right] \cdot \Delta,$$

$$\text{V} := \sup_{\theta \in \bar{\Theta}} \left| \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^{\theta}(t_k, s) \cdot \bar{J}^{\theta}(\bar{X}_s^{\theta}) ds \right| \leq \sqrt{\epsilon} \cdot \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |Z^{\theta}(t, 0) \cdot \bar{J}^{\theta}(\bar{X}_t^{\theta})| \right] \cdot \Delta.$$

By Lemma 4, there is a constant K such that for ϵ sufficiently small, $EI \leq K$. Similarly, there is a (perhaps larger) constant K such that for ϵ sufficiently small, $\text{II} + E\text{III} + \text{IV} + \text{V} \leq K\Delta$. The first statement, (32), follows easily.

(33) may be obtained in the same way by omitting the last term in (34). $\qquad \square$

LEMMA 10. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *There is a constant* $\tilde{K}$ *such that for all* $n > 0$, $k = 1, 2, \ldots, n$, *and* ϵ *(and hence also* $\delta = \delta(\epsilon)$*) sufficiently small,*

$$\text{(36)} \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} |F_k^{\varepsilon}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| \leq \Delta \cdot \tilde{K},$$

$$\text{(37)} \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} |\tilde{F}_k^{\varepsilon}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| \leq \Delta \cdot \tilde{K}.$$

Proof. We write

$$\text{(38)} \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} |F_k^{\varepsilon}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| \leq 2 \cdot \text{I} + 2 \cdot \text{II} + \text{III},$$

where

$$\text{I} := \sup_{\theta \in \bar{\Theta}} |\bar{X}_{t_k}^{\theta} - \bar{X}_{t_{k-1}}^{\theta}| \leq \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |(\nabla_x \bar{\lambda}^{\theta})(\bar{X}_t^{\theta})| \right] \cdot \Delta,$$

$$\text{II} := \sup_{\theta \in \bar{\Theta}} \left| 1_m - Z^{\theta}(t_k, t_{k-1}) \right| \cdot \sup_{\theta \in \bar{\Theta}} |\bar{X}_{t_{k-1}}^{\theta}|$$
$$\leq \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |Z^{\theta}(t, 0) \cdot (\nabla_x \bar{\lambda}^{\theta})(\bar{X}_t^{\theta})| \right] \cdot \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |\bar{X}_t^{\theta}| \right] \cdot \Delta,$$

$$\text{III} := \sup_{\theta \in \bar{\Theta}} \left| \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds \right| \leq \sqrt{\epsilon} \cdot \left[\sup_{\theta \in \bar{\Theta}, 0 \leq t \leq T} |Z^\theta(t, 0) \cdot \bar{J}^\theta(\bar{X}_t^\theta)| \right] \cdot \Delta.$$

The first statement, (36), follows easily. (37) may be obtained in the same way by omitting the last term in (38). □

LEMMA 11. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *There is a positive constant* $\tilde{K}$ *such that for all* $n > 0$ *and* $k = 1, 2, \ldots, n$,

$$\inf_{\theta \in \Theta} ||Q_k(\theta)|| \geq \Delta \cdot \tilde{K};$$

in particular, the inverse matrices Q_k^{-1} *exist and* $\sup_{\theta \in \Theta} ||Q_k^{-1}(\theta)|| \leq \Delta^{-1} \cdot \tilde{K}^{-1}$.

Proof. Recalling our nondegeneracy assumptions, $\bar{q}^\theta$ is uniformly nondegenerate by the argument of Theorem 11.3 in [24]. The result follows readily upon noting that the norms of the exponential terms in the definition of Q are bounded uniformly in all parameters away from zero. □

A.3. Lemmas on the contrast function.

LEMMA 12 (limit of the contrast function). *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *Fix* $n > 0$. *For any* $\eta > 0$,

$$\lim_{\epsilon \to 0} P \left(\sup_{\theta_1, \theta_2 \in \Theta} |U^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| > \eta \right) = 0.$$

Proof. We begin by writing

$$U^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) = \left[U^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) - U^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right]$$
$$+ \left[U^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right]$$
$$= [\Sigma_{k=1}^n A_k^\varepsilon(\theta_1, \theta_2)] + [\Sigma_{k=1}^n B_k^\varepsilon(\theta_1, \theta_2)],$$

where

$$A_k^\varepsilon(\theta_1, \theta_2) := \left[F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) - F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right]^T$$
$$\cdot Q_k^{-1}(\theta_2) \left[F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) + F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right],$$

$$B_k^\varepsilon(\theta_1, \theta_2) := \left[F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right]^T$$
$$\cdot Q_k^{-1}(\theta_2) \left[F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) + \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right].$$

As we have taken n to be fixed, it suffices to show that each term vanishes in probability uniformly in θ_1, θ_2 as $\epsilon \to 0$.

For A_k^ε, we have that

(39)
$$Q_k^{-1}(\theta_2) \left[F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) + F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) \right]$$

is bounded by Lemmas 9, 10, and 11 while

(40)
$$F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon, \theta_1}\}_{k=1}^n) - F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)$$
$$= [X_{t_k}^{\epsilon, \theta_1} - \bar{X}_{t_k}^{\theta_1}] - Z^{\theta_2}(t_k, t_{k-1}) \cdot [X_{t_{k-1}}^{\epsilon, \theta_1} - \bar{X}_{t_{k-1}}^{\theta_1}]$$

vanishes by Theorem 1 and the fact that the Frobenius norm of $Z^{\theta_2}(t_k, t_{k-1})$ is bounded.

For B_k^{ε}, we have that

$$(41) \qquad Q_k^{-1}(\theta_2)\left[F_k^{\varepsilon}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) + \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)\right]$$

is bounded by Lemmas 10 and 11 while

$$(42) \qquad F_k^{\varepsilon}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)$$
$$= -\sqrt{\epsilon}\int_{t_{k-1}}^{t_k} Z^{\theta_2}(t_k, s) \cdot \bar{J}^{\theta_2}(\bar{X}_s^{\theta_2})ds$$

vanishes by the vanishing prefactor $-\sqrt{\epsilon}$ and the fact that the Euclidean norm of $\int_{t_{k-1}}^{t_k} Z^{\theta_2}(t_k, s) \cdot \bar{J}^{\theta_2}(\bar{X}_s^{\theta_2})ds$ is bounded (indeed, proportionally to Δ). $\qquad\square$

LEMMA 13. *Assume Conditions 1 and 2 and, in the ∞ regime, Condition 3. There is a constant $\tilde{K}$ such that for any $\theta \in \Theta$ and $n > 0$,*

$$\sup_{\theta \in \Theta} \tilde{w}^{\theta}(\phi) \leq \tilde{K} \cdot \phi,$$

where

$$\tilde{w}^{\theta}(\phi) := \sup_{\theta_1, \theta_2 \in \Theta; |\theta_1 - \theta_2| \leq \phi} \left|\tilde{U}(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n)\right|.$$

Proof. We write

$$\tilde{U}(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) = \left[\Sigma_{k=1}^n C_k(\theta_1, \theta_2)\right] + \left[\Sigma_{k=1}^n D_k(\theta_1, \theta_2)\right],$$

where

$$C_k(\theta_1, \theta_2) := \left[\tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n)\right]^T$$
$$\cdot Q_k^{-1}(\theta_1)\left[\tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) + \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n)\right],$$
$$D_k(\theta_1, \theta_2) := \tilde{F}_k^T(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) \cdot \left[Q_k^{-1}(\theta_1) - Q_k^{-1}(\theta_2)\right] \cdot \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n)$$
$$= \tilde{F}_k^T(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) \cdot Q_k^{-1}(\theta_1) \cdot [Q_k(\theta_2) - Q_k(\theta_1)]$$
$$\cdot Q_k^{-1}(\theta_2) \cdot \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n).$$

It suffices to show that there is a constant K such that the absolute value of each term is bounded by $K \cdot \Delta \cdot |\theta_1 - \theta_2|$.

For C_k, the terms $Q_k^{-1}(\theta_1)[\tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) + \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n)]$ are bounded by Lemmas 10 and 11, so it will suffice to bound the terms

$$\left|\tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n)\right|,$$

individually, by a certain constant times $\Delta \cdot |\theta_1 - \theta_2|$.

Notice that

$$\tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^{\theta}\}_{k=1}^n) = \text{I} + \text{II} + \text{III},$$

where

$$\mathrm{I} := [\bar{X}_{t_k}^{\theta_2} - \bar{X}_{t_k}^{\theta_1}] - [\bar{X}_{t_{k-1}}^{\theta_2} - \bar{X}_{t_{k-1}}^{\theta_1}],$$

$$\mathrm{II} := \left[Z^{\theta_1}(t_k, t_{k-1}) - 1_m \right] \cdot \left[\bar{X}_{t_{k-1}}^{\theta_1} - \bar{X}_{t_{k-1}}^{\theta_2} \right],$$

$$\mathrm{III} := \left[Z^{\theta_1}(t_k, t_{k-1}) - Z^{\theta_2}(t_k, t_{k-1}) \right] \cdot \left[\bar{X}_{t_{k-1}}^{\theta_2} - \bar{X}_{t_{k-1}}^{\theta} \right].$$

By the triangle inequality,

$$\left| \tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) \right| \le |\mathrm{I}| + |\mathrm{II}| + |\mathrm{III}|.$$

First,

$$|\mathrm{I}| \le \int_{t_{k-1}}^{t_k} |\bar{\lambda}^{\theta_2}(\bar{X}_s^{\theta_2}) - \bar{\lambda}^{\theta_1}(\bar{X}_s^{\theta_1})| ds \le \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} |\nabla_\theta(\bar{\lambda}^{\tilde{\theta}}(X_t^{\tilde{\theta}}))| \right] \cdot \Delta \cdot |\theta_1 - \theta_2|,$$

second,

$$|\mathrm{II}| \le \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} |(\nabla_x \bar{\lambda}^{\tilde{\theta}})(\bar{X}_t^{\tilde{\theta}})| \right] \cdot \Delta \cdot \int_0^{t_{k-1}} |\bar{\lambda}^{\theta_1}(\bar{X}_s^{\theta_1}) - \bar{\lambda}^{\theta_2}(\bar{X}_s^{\theta_2})| ds$$

$$\le \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} |(\nabla_x \bar{\lambda}^{\tilde{\theta}})(\bar{X}_t^{\tilde{\theta}})| \right] \cdot \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} |\nabla_\theta(\bar{\lambda}^{\tilde{\theta}}(X_t^{\tilde{\theta}}))| \right] \cdot T \cdot \Delta \cdot |\theta_1 - \theta_2|;$$

third,

$$|\mathrm{III}| \le \left| Z^{\theta_1}(t_k, t_{k-1}) \right| \cdot \left| \int_{t_{k-1}}^{t_k} (\nabla_x \bar{\lambda}^{\theta_2})(\bar{X}_u^{\theta_2}) - (\nabla_x \bar{\lambda}^{\theta_1})(\bar{X}_u^{\theta_1}) du \right| \cdot \left| \bar{X}_{t_{k-1}}^{\theta_2} - \bar{X}_{t_{k-1}}^{\theta} \right|$$

$$\le \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} |Z^\theta(t, 0)| \right] \cdot \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} |\nabla_\theta \left((\nabla_x \bar{\lambda}^{\tilde{\theta}})(\bar{X}_t^{\tilde{\theta}}) \right) | \right]$$

$$\cdot \left[\sup_{\tilde{\theta} \in \Theta, 0 \le t \le T} 2|\bar{X}_t^{\tilde{\theta}}| \right] \cdot \Delta \cdot |\theta_1 - \theta_2|;$$

hence indeed there is a constant K such that

$$\left| \tilde{F}_k(\theta_1; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) - \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^\theta\}_{k=1}^n) \right| \le K \cdot \Delta \cdot |\theta_1 - \theta_2|.$$

For D_k, the terms $Q_k^{-1}(\theta_i) \cdot \tilde{F}_k(\theta_2; \{\bar{X}_{t_k}^\theta\}_{k=1}^n)$ are bounded by Lemmas 10 and 11, so using the fact that the Frobenius norm is equivalent to the operator norm, it will suffice to bound the terms $|Q_k(\theta_2) - Q_k(\theta_1)|$, individually, by a certain constant times $\Delta \cdot |\theta_1 - \theta_2|$.

We have

$$Q_k(\theta_2) - Q_k(\theta_1) = \mathrm{IV} + \mathrm{V},$$

where

$$\mathrm{IV} := \int_{t_{k-1}}^{t_k} Z^{\theta_2}(t_k, s) \cdot \left(\bar{q}^{\theta_2}(\bar{X}_s^{\theta_2}) - \bar{q}^{\theta_1}(\bar{X}_s^{\theta_1}) \right) \cdot e^{\int_s^{t_k} (\nabla_x \bar{\lambda}^{\theta_2})^T (\bar{X}_u^{\theta_2}) du} ds,$$

$$\mathrm{V} := \int_{t_{k-1}}^{t_k} \left(Z^{\theta_2}(t_k, s) - Z^{\theta_1}(t_k, s) \right) \cdot \bar{q}^{\theta_1}(\bar{X}_s^{\theta_1}) \cdot \left(Z^{\theta_2}(t_k, s) + Z^{\theta_1}(t_k, s) \right);$$

by the triangle inequality,

$$|Q_k(\theta_2) - Q_k(\theta_1)| \leq |\mathrm{IV}| + |\mathrm{V}|.$$

First,

$$|\mathrm{IV}| \leq \left[\sup_{\tilde{\theta} \in \Theta, 0 \leq t \leq T} \left| Z^{\tilde{\theta}}(t, 0)(Z^{\tilde{\theta}})^T(t, 0) \right| \right] \cdot \left[\sup_{\tilde{\theta} \in \Theta, 0 \leq t \leq T} || \nabla_\theta \left(\bar{q}^{\tilde{\theta}}(\bar{X}_t^{\tilde{\theta}}) \right) || \right] \cdot \Delta \cdot |\theta_1 - \theta_2|;$$

second,

$$|\mathrm{V}| \leq \left[\sup_{\tilde{\theta} \in \Theta, 0 \leq t \leq T} |Z^\theta(t, 0)| \right] \cdot \left[\sup_{\tilde{\theta} \in \Theta, 0 \leq t \leq T} |\nabla_\theta \left((\nabla_x \bar{\lambda}^{\tilde{\theta}})(\bar{X}_t^{\tilde{\theta}}) \right) | \right]$$

$$\cdot \left[\sup_{\tilde{\theta} \in \Theta, 0 \leq t \leq T} ||\bar{q}^{\tilde{\theta}}(\bar{X}_t^{\tilde{\theta}})|| \right] \cdot \Delta \cdot |\theta_1 - \theta_2|,$$

where the difference of exponentials has been handled exactly as in the bound above for $|\mathrm{III}|$; hence indeed there is a (perhaps larger) constant K such that

$$|Q_k(\theta_2) - Q_k(\theta_1)| \leq K \cdot \Delta \cdot |\theta_1 - \theta_2|,$$

completing the proof of the lemma. $\qquad\square$

A.4. Lemmas for frequent sampling. The contrast functions upon which our estimators are founded are defined as sums over a number of indices equal to the number of sampled data n. When n is fixed, termwise convergence as $\epsilon \to 0$ is enough to deduce convergence of the sums. When, however, n is permitted to increase simultaneously as $\epsilon \to 0$, one must examine the convergence of sums over increasingly many indices. Convergence is easily deduced if statements of termwise convergence can be strengthened so that they uniformly outpace the increasing number of indices. The strengthened bounds that allow us to take this approach are established in this section.

LEMMA 14. *Assume Conditions* 1 *and* 2 *and, in the* ∞ *regime, Condition* 3. *Assume that the sampling interval* $\Delta := T/n$ *does not decrease too quickly relative to* ϵ; *that is, suppose that there is a finite constant* G *such that always* $\epsilon \cdot n \leq G$. *There is a constant* $\tilde{K}$ *such that for* ϵ *(and hence also* $\delta = \delta(\epsilon)$) *sufficiently small, independently of* (n, k),

$$(43) \qquad E|F_k^\varepsilon(\theta; \{X_{t_k}^{\epsilon, \theta}\}_{k=1}^n)|^2 \leq \tilde{K} \cdot \epsilon \cdot \Delta,$$

$$(44) \qquad E|\tilde{F}_k(\theta; \{X_{t_k}^{\epsilon, \theta}\}_{k=1}^n)|^2 \leq \tilde{K} \cdot \epsilon \cdot \Delta.$$

Proof. Recall that

$$(45) \qquad F_k^\varepsilon(\theta; \{X_{t_k}^{\epsilon, \theta}\}_{k=1}^n) = \left[[X_{t_k}^{\epsilon, \theta} - \bar{X}_{t_k}^\theta] - Z^\theta(t_k, t_{k-1}) \cdot [X_{t_{k-1}}^{\epsilon, \theta} - \bar{X}_{t_{k-1}}^\theta] \right]$$

$$- \sqrt{\epsilon} \int_{t_{k-1}}^{t_k} Z^\theta(t_k, s) \cdot \bar{J}^\theta(\bar{X}_s^\theta) ds$$

$$= \mathrm{I} + \mathrm{II} + \mathrm{III},$$

where

$$\text{I} := \int_{t_{k-1}}^{t_k} \left[\frac{\epsilon}{\delta} b(X_s^{\varepsilon,\theta}, Y_s^{\epsilon,\theta}) + c(X_s^{\varepsilon,\theta}, Y_s^{\epsilon,\theta}) - \bar{\lambda}^\theta(\bar{X}_s^\theta) \right] ds,$$

$$\text{II} := -\sqrt{\epsilon} \int_{t_{k-1}}^{t_k} \sigma(X_s^{\varepsilon,\theta}, Y_s^{\epsilon,\theta}) dW_s,$$

$$\text{III} := \left(1 - Z^\theta(t_k, t_{k-1}) \right) \cdot [X_{t_{k-1}}^{\varepsilon,\theta} - \bar{X}_{t_{k-1}}^\theta];$$

it suffices to consider each of these separately.

I is handled by arguments similar to those presented in the proofs of Theorem 2 and Lemma 4. It follows by the calculations (11) and, in the ∞ regime, (29), that there is a constant K, which may be chosen independently of (n, k), such that

$$E|\text{I}|^2 \leq \delta^2 \cdot K \leq \epsilon \cdot \Delta \cdot K,$$

where we have used the fact that δ is $O(\epsilon)$ and the assumption that ϵ is $O(\Delta)$.

Meanwhile, it is clear that there is a (perhaps larger) constant K such that

$$E|\text{II}|^2 \leq \epsilon \cdot \Delta \cdot K.$$

Finally, in light of Theorem 1 and the fact that $||1_m - Z^{\theta_2}(t_k, t_{k-1})||$ is bounded by a finite constant times Δ as $\Delta \to 0$, it is clear that for a (perhaps larger) constant K,

$$E|\text{III}|^2 \leq \epsilon \cdot \Delta^2 \cdot K \leq \epsilon \cdot \Delta \cdot T \cdot K.$$

The gives the first statement, (43), in the ∞ regime. The proof for the γ regime is similar and easier.

(44) may be obtained in the same way by omitting the last term in (45). $\qquad \square$

LEMMA 15. *Assume Conditions* 1 *and* 2 *and, in the ∞ regime, Condition* 3. *Assume that the sampling interval $\Delta := T/n$ does not decrease too quickly relative to ϵ; that is, suppose that there is a finite constant G such that always $\epsilon \cdot n \leq G$. There is a constant $\tilde{K}$ such that for ϵ (and hence also $\delta = \delta(\epsilon)$) sufficiently small, independently of (n, k),*

$$(46) \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} E|F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)| \leq \Delta \cdot \tilde{K},$$

$$(47) \qquad \sup_{(\theta_1, \theta_2) \in \bar{\Theta}^2} E|\tilde{F}_k(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n)| \leq \Delta \cdot \tilde{K}.$$

Proof. Let us examine the proof of Lemma 9, with particular attention to (35). It is not difficult to see that a constant K may be found such that for ϵ sufficiently small,

$$EI \leq K(\sqrt{\epsilon} + \sqrt{\Delta})\sqrt{\Delta}$$
$$\leq K(\sqrt{\epsilon \cdot n} + 1)\Delta.$$

Thus one sees that with the new assumption that $\epsilon \cdot n \leq G$, EI is bounded for ϵ sufficiently small, independently of n, by a finite constant times Δ. The same having been noted for $\text{II} + E\text{III} + \text{IV} + \text{V}$, the desired bounds follow. $\qquad \square$

LEMMA 16 (limit of the contrast function as $\epsilon + \Delta \to 0$). *Assume Conditions* 1 *and* 2 *and, in the ∞ regime, Condition* 3. *Assume that the sampling interval $\Delta := T/n$ does not decrease too quickly relative to ϵ, so that ϵ is $o(\Delta)$ as $\Delta \to 0$. For any $\eta > 0$,*

$$\lim_{(\epsilon+\Delta)\to 0} P\left(\sup_{\theta_1,\theta_2 \in \Theta} |U^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n) - \tilde{U}(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| > \eta\right) = 0.$$

Proof. Let us examine the proof of Lemma 12. It will suffice to show that the terms A_k^ε, B_k^ε vanish *faster than* Δ in probability uniformly in θ_1, θ_2 as $\epsilon + \Delta \to 0$. Recall that Lemmas 10 and 11 apply uniformly in n. Thus, (41) is bounded uniformly in n and, with the new assumption that $\epsilon \cdot n \leq G$, one may replace Lemma 9 with Lemma 15 to establish the same for (39). Meanwhile, it is clear that (43) vanishes faster than Δ. Thus, if we establish that (41) also vanishes faster than Δ, we will have proven the lemma.

Splitting the right-hand side of (41), we obtain

$$F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n) - F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n) = \text{I} + \text{II} + \text{III},$$

where

$$\text{I} := \int_{t_{k-1}}^{t_k} \left[\frac{\epsilon}{\delta}b(X_s^{\epsilon,\theta_1}, Y_s^{\epsilon,\theta_1}) + c(X_s^{\epsilon,\theta_1}, Y_s^{\epsilon,\theta_1}) - \bar{\lambda}^{\theta_1}(\bar{X}_s^{\theta_1})\right] ds,$$

$$\text{II} := \sqrt{\epsilon}\int_{t_{k-1}}^{t_k} \sigma(X_s^{\epsilon,\theta_1}, Y_s^{\epsilon,\theta_1})dW_s,$$

$$\text{III} := \left(1_m - Z^{\theta_2}(t_k, t_{k-1})\right) \cdot [X_{t_{k-1}}^{\epsilon,\theta_1} - \bar{X}_{t_{k-1}}^{\theta_1}].$$

I is handled by an argument similar to the proof of Lemma 4—letting χ be as in (3), one applies the Itô formula to $\chi(X_t^\varepsilon, Y_t^\varepsilon)$ and rearranges the terms to see that there is a constant K, which may be chosen independently of (n, k), such that

$$E|\text{I}| \leq \delta \cdot K \leq \epsilon \cdot K,$$

where we have used the fact that δ is $O(\epsilon)$. Meanwhile, it is clear that there is likewise a (perhaps larger) constant K such that

$$E|\text{II}| \leq \sqrt{\epsilon\Delta} \cdot K.$$

Finally, in light of Theorem 1 and the fact that $\left(1_m - Z^{\theta_2}(t_k, t_{k-1})\right)$ is bounded by a finite constant times Δ, it is clear that for a (perhaps larger) constant K,

$$E|\text{III}| \leq \sqrt{\epsilon}\Delta \cdot K.$$

Hence,

$$E|F_k^\varepsilon(\theta_2; \{X_{t_k}^{\epsilon,\theta_1}\}_{k=1}^n) - F_k^\varepsilon(\theta_2; \{\bar{X}_{t_k}^{\theta_1}\}_{k=1}^n)| \leq (\epsilon + \sqrt{\epsilon\Delta}) \cdot 2^3 \cdot K,$$

which is enough since we have assumed ϵ is $o(\Delta)$. □

REFERENCES

[1] R. Azencott, A. Beri, and I. Timofeyev, *Adaptive sub-sampling for parametric estimation of Gaussian diffusions*, J. Stat. Phys., 139 (2010), pp. 1066–1089.

[2] R. Azencott, A. Beri, A. Jain, and I. Timofeyev, *Sub-sampling and parametric estimation for multiscale dynamics*, Commun Math. Sci., 11 (2013), pp. 939–970.

[3] J. P. N. Bishwal, *Parameter Estimation in Stochastic Differential Equations*, Lecture Notes in Math. 1923, Springer, Berlin, 2008.

[4] A. Chauvière, L. Preziosi, and C. Verdier, *Cell Mechanics: From Single Scale-Based Models to Multiscale Modeling*, CRC, Boca Raton, FL, 2010.

[5] C. Chicone, *Ordinary Differential Equations with Applications*, 2nd ed., Springer, New York, 2006.

[6] D. Dacunha-Castelle and M. Duflo, *Probability and Statistics*, Vol. 2, Springer, 1986.

[7] J. Feng, M. Forde, and J.-P. Fouque, *Short-maturity asymptotics for a fast mean-reverting Heston stochastic volatility model*, SIAM J. Financial Math., 1 (2010), pp. 126–141.

[8] J. Feng, J.-P. Fouque, and R. Kumar, *Small-time asymptotics for fast mean-reverting stochastic volatility models*, Ann. Appl. Probab., 22 (2012), pp. 1541–1575.

[9] J.-P. Fouque, G. Papanicolaou, and K. R. Sircar, *Derivatives in Financial Markets with Stochastic Volatility*, Cambridge University Press, 2000.

[10] S. Gailus and K. Spiliopoulos, *Statistical inference for perturbed multiscale dynamical systems*, Stochastic Process. Appl., 127 (2016), pp. 419–448.

[11] D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, New York, 2001.

[12] R. Guy, C. Larédo, and E. Vergu, *Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient*, Stochastic Process. Appl., 124 (2014), pp. 51–80.

[13] W. Janke, *Rugged Free-Energy Landscapes*, Lecture Notes in Phys. 736, Springer, Berlin, 2008.

[14] V. K. Jirsa, W. C. Stacey, P. P. Quilichini, A. I. Ivanov, and C. Bernard, *On the nature of seizure dynamics*, Brain, 137 (2014), pp. 2210–2230.

[15] S. Krumscheid, G. A. Pavliotis, and S. Kalliadasis, *Semiparametric drift and diffusion estimation for multiscale diffusions*, Multiscale Model. Simul., 11 (2013), pp. 442–473.

[16] Y. Kutoyants, *Identification of Dynamical Systems with Small Noise*, Kluwer Academic, Dordrecht, The Netherlands, 1994.

[17] Y. Kutoyants, *Parameter Estimation for Stochastic Processes*, Helderman, Berlin, 1984.

[18] Y. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*, Springer, London, 2004.

[19] A. J. Majda, C. Franzke, and B. Khouider, *An applied mathematics perspective on stochastic modelling for climate*, Philos. Trans. Roy. Soc. A, 366 (2008), pp. 2429–2455.

[20] A. Papavasiliou, G. A. Pavliotis, and A. M. Stuart, *Maximum likelihood drift estimation for multiscale diffusions*, Stochastic Process. Appl., 119 (2009), pp. 3173–3210.

[21] E. Pardoux and A. Y. Veretennikov, *On the Poisson equation and diffusion approximation 1*, Ann. Probab., 29 (2001), pp. 1061–1085.

[22] E. Pardoux and A. Y. Veretennikov, *On Poisson equation and diffusion approximation 2*, Ann. Probab., 31 (2003), pp. 1166–1192.

[23] G. A. Pavliotis and A. M. Stuart, *Parameter estimation for multiscale diffusions*, J. Stat. Phys., 127 (2007), pp. 741–781.

[24] G. A. Pavliotis and A. M. Stuart, *Multiscale methods: Averaging and homogenization*, Texts Appl. Math. 53, Springer, New York, 2008.

[25] B. L. S. P. Rao, *Statistical Inference for Diffusion Type Processes*, Arnold, London, 1999.

[26] K. Spiliopoulos, *Fluctuation analysis and short time asymptotics for multiple scales diffusion processes*, Stoch. Dyn., 14 (2014), 1350026.

[27] K. Spiliopoulos and A. Chronopoulou, *Maximum likelihood estimation for small noise multiscale diffusions*, Stat. Inference Stoch. Process., 16 (2013), pp. 237–266.

[28] M. Sørensen and M. Uchida, *Small diffusion asymtptotics for discretely sampled stochastic differential equations*, Bernoulli, 9 (2003), pp. 1051–1069.

[29] M. Uchida, *Estimation for discretely observed small diffusions based on approximate martingale estimating functions*, Scand. J. Stat., 31 (2004), pp. 553–566.

[30] A. Y. Veretennikov, *On polynomial mixing bounds for stochastic differential equations*, Stochastic Process. Appl., 70 (1997), pp. 115–127.

[31] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia, *A tale of two time scales*, J. Amer. Statist. Assoc., 100 (2005), pp. 1394–1411.

[32] R. Zwanzig, *Diffusion in a rough potential*, Proc. Natl. Acad. Sci. USA, 85 (1988), pp. 2029–2030.