Parametric Estimation from Approximate Data: Non-Gaussian Diffusions

Robert Azencott¹ · Peng Ren¹ · Ilya Timofeyev¹

Received: 30 July 2015 / Accepted: 18 September 2015 / Published online: 1 October 2015
© Springer Science+Business Media New York 2015

Abstract We study the problem of parameters estimation in indirect observability contexts, where $X_t \in \mathbb{R}^r$ is an unobservable stationary process parametrized by a vector of unknown parameters and all observable data are generated by an approximating process $Y_t^\varepsilon$ which is close to $X_t$ in $L^4$ norm. We construct consistent parameter estimators which are smooth functions of the sub-sampled empirical mean and empirical lagged covariance matrices computed from the observable data. We derive explicit optimal sub-sampling schemes specifying the best paired choices of sub-sampling time-step and number of observations. We show that these choices ensure that our parameter estimators reach optimized asymptotic $L^2$-convergence rates, which are constant multiples of the $L^4$ norm $||Y_t^\varepsilon - X_t||$.

Keywords Parametric estimation · Non-Gaussian diffusions · Empirical covariance estimators · Indirect observability

1 Introduction

The amount of available observational data has increased massively in recent years due to rapid technological advances in science and engineering. Often, it is desirable to fit an appropriate parametrized stochastic model to the available data and then use this model for forecasting, analysis, etc. Stochastic processes $X_t$ driven by systems of stochastic differential equations (SDEs) have often been used for this purpose so that both parametric and non-parametric techniques for fitting SDEs to the available data have a rich history.
SDE parameters by classes of estimators depending only on main mathematical goal in this indirect observability context is to estimate the underlying conditions \( \phi(\cdot) \), but for Gaussian limiting processes \( X_t \) generate consistent estimators \( \hat{\theta} \) of the approximate data, a natural approach is to first derive "ideal" estimators as specific functionals of observed trajectories \( Y^\varepsilon_t \); of course as \( \varepsilon \to 0 \), one wants these estimators to be consistent and approximately efficient. Recent references for consistent estimation of parametrized diffusions under indirect observability include [4,6,7,21,24,25,35,55,56]. In financial applications, indirect observability situations emerge as soon as dynamic models involve the (unobservable) volatilities of assets, and their replacement by observable approximations of volatilities, such as "realized volatilities" computed from stock prices (see, e.g., [11–13,15,18,34]). Indirect observability is also a natural context for stock prices dynamics based on precise noise microstructure (see, e.g., [2,70]).

Our indirect observability framework covers a broad range of SDE systems driving unobservable multidimensional processes \( X_t \), where the drift and diffusion terms are fairly generic smooth functions of an unknown parameter vector \( \theta \). Such SDEs are widely used in engineering and automatic control, population evolution, atmospheric and ocean dynamics, stock prices dynamics, options pricing, etc. to approximate the behavior of the leading variables of interest. To construct consistent estimators \( \hat{\theta} \) of \( \theta \) based on observed trajectories \( Y^\varepsilon_{[0,t]} \) of the approximate data, a natural approach is to first derive "ideal" estimators as specific functionals \( \phi(X_{[0,t]}) \) of the unobservable trajectory \( X_{[0,t]} \), and then prove that under adequate conditions \( \phi(Y^\varepsilon_{[0,t]}) \to \theta \) as \( \varepsilon \to 0 \) and \( N(\varepsilon) \to \infty \), where \( N(\varepsilon) \) is the size of the observational data sample. In [4,6,7], we had combined this approach with data sub-sampling to generate consistent estimators \( \hat{\theta} \) based on approximate observable data \( Y^\varepsilon_t \) with fairly generic joint distributions, but for Gaussian limiting processes \( X_t \), and we had also determined how nearly optimal sub-sampling rates should depend on \( \varepsilon \). In this paper we extend our indirect observability analysis to stationary processes \( X_t \) such that \( Y^\varepsilon_t \to X_t \) in \( L^4 \) as \( \varepsilon \to 0 \), but with weak restrictions on the joint distributions of \( X_t \) and \( Y^\varepsilon_t \), which can both be non-Gaussian.

Discretized sub-sampling is standard to collect data from a continuous process. The observable data are then of the form \( Y^\varepsilon_{nh} \), where the observational time-step \( \delta \) is determined by

(see [8,39,46,61] for a general overview). Non-parametric approaches for SDE data modeling have used Bayesian methods as in [27,53,54,59,68], exploited the spectral properties of the infinitesimal generator [20,21,37], or have developed maximum likelihood function estimation as in [44,45,65], as well as drift and diffusion estimates by conditional expectations of process dynamics over short time intervals [14,19,31,41,64,67], with potential use of kernel based techniques as in [10,69]. For parametrized SDE models, various moments based parameter estimators (see [32] and references therein) have been implemented, as well as approximate maximum-likelihood parameters estimators after time discretization of the SDEs (see for instance [1,9,17,58]). Minimum-contrast estimators have also been used for parametric estimation of diffusions [22,35]. The asymptotic consistency and efficiency of SDEs model fitting to data have been well analyzed in the literature, although computational issues remain key questions in high dimensional applications.

In this paper we address the problem of parameter estimation for multi-dimensional diffusions in a specific context, namely in situations when the process \( X_t \) to be modeled is not directly observable (i.e. \( X_t \) is unobservable). Instead, the observed data are generated by an approximating process \( Y^\varepsilon_t \) which involves a small “scaling” parameter \( \varepsilon \), and the SDEs driving \( X_t \) are discovered by asymptotic analysis as \( \varepsilon \to 0 \). Moreover, often the precise dynamics of the process \( Y^\varepsilon_t \) is not known or too complex to be explicitly formalized. In such indirect observability contexts, where the SDE system driving the unobservable \( X_t \) is parametrized by an unknown vector \( \theta \), but the observable \( Y^\varepsilon_t \) is generated by unknown dynamics approximating the behavior of \( X_t \), a natural goal is to efficiently estimate \( \theta \) from the observable data \( Y^\varepsilon_t \), assuming that for some adequate norm \( \| Y^\varepsilon_t \| \to \| X_t \| \) as \( \varepsilon \to 0 \). The main mathematical goal in this indirect observability context is to estimate the underlying SDE parameters by classes of estimators depending only on \( \varepsilon \) and on the observable data \( Y^\varepsilon_t \); of course as \( \varepsilon \to 0 \), one wants these estimators to be consistent and approximately efficient. Recent references for consistent estimation of parametrized diffusions under indirect observability include [4,6,7,21,24,25,35,55,56]. In financial applications, indirect observability situations emerge as soon as dynamic models involve the (unobservable) volatilities of assets, and their replacement by observable approximations of volatilities, such as “realized volatilities” computed from stock prices (see, e.g., [11–13,15,18,34]). Indirect observability is also a natural context for stock prices dynamics based on precise noise microstructure (see, e.g., [2,70]).
data acquisition protocols for sensors recordings or by the computational time-step for observables generated by numerical PDE models. As is well known, adequate data sub-sampling can reduce computational overhead without sacrificing estimators accuracy. So we introduce a user selected subsampling time-step $\Delta(\varepsilon)$, which will be a multiple of $\delta$ and will fix the observational sample $\{Y_{n\Delta(\varepsilon)}^\varepsilon, n = 1, \ldots, N(\varepsilon)\}$ retained to estimate $\theta$. We then address the key issue of how to optimize parameter estimators performance by seeking the “best” asymptotic choices for the number of observations $N(\varepsilon)$ and the sub-sampling time-step $\Delta(\varepsilon)$. We thus derive explicit relations between $\Delta(\varepsilon)$, $N(\varepsilon)$, and $\rho(\varepsilon) = \|Y_t^\varepsilon - X_t\|_4$ ensuring nearly optimal behavior for parameter estimation errors as $\varepsilon \to 0$. This is achieved by focusing on estimators which are (not necessarily explicit) smooth functions of empirical lagged moments up to order two computed from the $N(\varepsilon)$ sub-sampled observables $\{Y_{n\Delta(\varepsilon)}^\varepsilon, n = 1, \ldots, N(\varepsilon)\}$.

Since in practice the available numbers of observations remain moderate, our nearly optimal choices for the pair $(N(\varepsilon), \Delta(\varepsilon))$ are constructed to simultaneously minimize the size of parameter estimation errors and the computational/observational complexity.

The paper is organized as follows. Basic assumptions about our indirect observability setup are given in Sect. 2. Our main results about speed of convergence for parameters estimators based on observable data and the associated characterization of optimal sub-sampling schemes are presented in Sect. 5. In Sect. 3 we outline the main class of parameter estimators studied here, namely smooth functions of lagged moments of order $\leq 2$. Section 4 contains key technical results on $L^2$ consistency of moments estimators computed from unobservable data. Potential applications of our results to stationary multi-dimensional diffusions and to Heston SDEs are discussed in the two Sects. 2.2 and 6.

2 Indirect Observability Setup

2.1 Basic Indirect Observability Hypotheses

2.1.1 Notations

For any matrix $M = (M_{i,j})$, we set $\|M\| = \sup_{i,j} M_{i,j}$, and $M^*$ is the transpose of $M$. The $L^p$ norm of a random matrix $M$ is denoted by $\|M\|_p = \mathbb{E}(\|M\|^p)^{1/p}$.

2.1.2 Indirect Observability

Our formal indirect observability setup $X_t = \lim_{\varepsilon \to 0} Y_t^\varepsilon$ involves

- a set of directly observable continuous time stochastic processes $Y_t^\varepsilon \in \mathbb{R}^r$ indexed by a small “scale” parameter $\varepsilon > 0$, with $L^4$ norms $\|Y_t^\varepsilon\|_4$ uniformly bounded for all $\varepsilon > 0$ and $t \geq 0$,
- an unobservable strictly stationary process $X_t = \lim_{\varepsilon \to 0} Y_t^\varepsilon$, where convergence holds at uniform $L^4$-speed $\rho(\varepsilon) \to 0$ as $\varepsilon \to 0$, so that
  \[ \|Y_t^\varepsilon - X_t\|_4 \leq \rho(\varepsilon), \text{ for all } t \geq 0, \varepsilon > 0. \]  

Moreover, we will assume that

- the mean $\mu$ of $X_t$ and the lagged covariance matrices $K(u) = \mathbb{E}(X_t X_{t+u}^*) - \mu \mu^*$ are $C^1$ functions of an unknown parameter vector $\theta \in \Theta$, with $\Theta$ open in $\mathbb{R}^p$,
- as functions of the time lag $u$, the matrices $K(u)$ are locally Lipschitz, uniformly in $\theta$.  

(Springer)
Since $X_t$ is not directly observable, our main goal here is to construct observable estimators $\theta^\varepsilon$ of $\theta$, i.e. estimators depending only on the observable data $Y^\varepsilon_t$, and hence of the form

$$\theta^\varepsilon = \phi^\varepsilon(Y^\varepsilon_1, \ldots, Y^\varepsilon_t),$$

where the number $q$ of observables and the instants $t_1, \ldots, t_q$ depend on $\varepsilon$, and each Borel function $\phi^\varepsilon : \mathbb{R}^q \to \mathbb{R}^m$ is deterministic. As $\varepsilon \to 0$, achieving consistency in probability for the estimators $\theta^\varepsilon$ will require to specify nearly optimal choices for $q(\varepsilon)$ and for the time grid $t_1(\varepsilon), \ldots, t_q(\varepsilon)$, and to clarify how these choices depend on the approximation speed $\rho(\varepsilon)$.

### 2.2 Multi-dimensional Diffusions Under Indirect Observability

#### 2.2.1 Stationary Multi-dimensional Diffusions

Many practical examples of indirect observability are linked to the approximation of high dimensional multiscale systems by reduced stochastic differential equations obtained by averaging and homogenization [4,21,42,50,51,55–57]. For instance this is the case of homogenization applications to the atmosphere-ocean science [26,28,29,48,49]. In those contexts the fully explicit mathematical dynamics for the observable $Y^\varepsilon_t$ is typically too complex to be fully modeled. However, the unobservable limit process $X_t \in \mathbb{R}^r$ is often modeled by a relatively low-dimensional strictly stationary multi-dimensional diffusion driven by an SDE system with coefficients depending smoothly on the parameter vector $\theta \in \Theta \subset \mathbb{R}^p$.

These SDE systems, derived by an ad-hoc analysis of the main “mechanisms” generating the $Y^\varepsilon_t$ observables, are of the form

$$dX_t = b(X_t, \theta)dt + \sigma(X_t, \theta)dW_t,$$

where $W_t$ is a multi-dimensional Brownian motion, the drift $b(x, \theta)$ and the matrix $\sigma(x, \theta)$ are $C^\infty$ functions of $x \in \mathbb{R}^r$ and $\theta \in \mathbb{R}^p$, the matrix $a(x, \theta) = \frac{1}{2} \sigma \sigma^*$ is invertible for all $x$ and $\theta$.

The transition density $p_\theta(t, x, y)$ of $X_t$ then depends smoothly on $\theta$ [66] and is the fundamental solution of a parabolic PDE with coefficients $\frac{1}{2}a$ and $b$. The literature (see for instance [30,40,47]) has extensively discussed these Fokker–Planck PDEs, as well as the elliptic PDE verified by the stationary density $q_\theta(x, y) = \lim_{t \to \infty} p_\theta(t, x, y)$, when $X_t$ is strictly stationary. In dimension 1, strict stationarity of $X_t$ is equivalent to integrability in $x \in \mathbb{R}$ of $p(x) = \frac{1}{a(x)} \exp(b(x)a(x))$ and (see [43]) the stationary density of $X_t$ is then proportional to $p(x)$. In higher dimensions, the literature does not seem to provide easy to use sufficient conditions for stationarity, so that stationarity has to be verified in each specific application.

#### 2.2.2 Parameter Estimation for Diffusions

The abundant literature on parameter estimation for multi-dimensional diffusions such as $X_t$ (see overviews [39,46,61] and references therein) offers a broad range of parameter estimators which are non explicit but numerically computable functions

$$\hat{\theta} = g_q(X_{t_1}, \ldots, X_{t_q})$$

of q diffusion “data” $X_{t_1}, \ldots, X_{t_q}$. Estimators of this type can for instance be numerically derived by approximate maximum likelihood after time discretization (see, e.g., [1,9,27,33,39,46,58,61]).

Under variously formulated sufficient conditions on the diffusion, maximum likelihood estimators of $\theta$ become asymptotically consistent for $q$ sufficiently large and for dense enough
specific time grids $t_1, \ldots, t_q$. However, in our indirect observability setup, these estimators are obviously not “observable”, since they involve the non-observable diffusion data $X_{t_1}, \ldots, X_{t_q}$. When “good” choices are known for the $g_q$ functions in (4), one can naturally attempt to construct consistent observable estimators $\theta^\varepsilon$ by setting

$$\theta^\varepsilon = g_q(Y^\varepsilon_{t_1}, \ldots, Y^\varepsilon_{t_q})$$

with “adequate” choices for the number of observations $q(\varepsilon)$ and for the time grid $t_1(\varepsilon), \ldots, t_q(\varepsilon)$. A key technical goal in our paper is to “optimally” select $q(\varepsilon)$ and this time grid as functions of $\varepsilon$, and to link such nearly optimal sub-sampling schemes to the $L^2$-speed of approximation $\rho(\varepsilon)$ of $X_t$ by the observable $Y^\varepsilon_t$.

2.2.3 Sparsely Parametrized Stationary Diffusions

For “uniformly elliptic” stationary diffusions $X_t$ driven by Eq. (3), the matrix $a$, its inverse $a^{-1}$, and the drift $b$ are classically assumed to be uniformly bounded for $x \in \mathbb{R}^r$ and $\theta \in \mathbb{R}^p$. The well known Aronson bounds of the diffusion transition density $p_\theta(t, x, y)$ (see [3]) then imply the finiteness of all lagged moments of arbitrary order for $X_t$, and due to the smoothness of $p_\theta$, all these lagged moments are then necessarily smooth functions of $\theta$.

2.2.4 Conjecture

On the basis of multiple concrete examples of stationary diffusions, we conjecture that when the matrices of diffusion coefficients $a$ and $b$ in Eq. (3) are analytic functions of the $p$-dimensional parameter vector $\theta$, one can then find a vector $\Psi = [\Psi_1, \ldots, \Psi_p]$ of $p$ lagged moments of order $\leq 2$ of $X_t$, which uniquely determine $\theta$ as a (non explicit) smooth function $\theta = G(\Psi)$ of $\Psi$. For such “sparsely parametrized” diffusions, parameter estimators based on estimation of lagged moments become natural targets, as outlined in the next section.

3 Parameter Estimators Based on Lagged Moments of Order $\leq 2$

3.1 Sparsely Parametrized Stationary Processes

The preceding conjecture on multidimensional stationary diffusion processes leads us to introduce the following definition.

**Definition 1** We say that a strictly stationary process $X_t$ is sparsely parametrized by $\theta \in \mathbb{R}^p$ whenever one can find a vector $\Psi = [\Psi_1, \ldots, \Psi_p]$ of $p$ lagged moments of order $\leq 2$ of $X_t$, which uniquely determine $\theta$ as a smooth function $\theta = G(\Psi)$ of $\Psi$.

Let $X_t$ be a sparsely parametrized stationary process embedded in an indirect observability setup $X_t = \lim_{\varepsilon \to 0} Y^\varepsilon_t$. As $\varepsilon \to 0$, constructing consistent observable estimators for all lagged moments of order $\leq 2$ of $X_t$ is then clearly equivalent to constructing consistent observable estimators $\hat{\theta}^\varepsilon$ for $\theta$. Indeed if $\theta = G(\Psi)$, one can simply set $\hat{\theta}^\varepsilon = G(\hat{\Psi}^\varepsilon)$, where $\hat{\Psi}^\varepsilon$ is an observable estimator of the vector of moments $\Psi$. Observable estimators for the lagged 2nd order moments $\Psi_j$ are naturally provided by the empirical lagged covariances of $Y^\varepsilon_t$, and hence consistent observable estimators of $\theta$ can then be constructed as smooth functions of empirical lagged moments of order $\leq 2$ of the observable data $Y^\varepsilon_t$. 
3.2 Sub-sampled Empirical Moments

As pointed out in many papers on indirect observability (see, e.g., [4,21,34,55,56]), correctly scaled sub-sampling of observable data can reduce computational overhead and decrease the bias of parameter estimators. A main technical point here will then be to select nearly optimal functions of $\varepsilon$ for the sub-sampling time step $\Delta(\varepsilon)$ of observable data as well as for the overall number $N(\varepsilon)$ of sub-sampled observables. So we now define adaptive sub-sampling schemes for empirical estimators of lagged covariances.

**Definition 2** Adaptive sub-sampling schemes will be specified by two functions of $\varepsilon$: the sub-sampling time step $\Delta(\varepsilon)$ and the number $N(\varepsilon)$ of sub-sampled observations, and we will impose the natural conditions

$$\Delta(\varepsilon) \to 0 \quad \text{and} \quad N(\varepsilon) \Delta(\varepsilon) \to \infty \quad \text{as} \quad \varepsilon \to 0. \tag{5}$$

Fix an adaptive sub-sampling scheme $\Delta(\varepsilon), N(\varepsilon)$. Each time lag $u$ will then be approximated by $\kappa$ with the integer $\kappa$ given by

$$\kappa = \kappa(u, \varepsilon) = \left\lfloor \frac{u}{\Delta(\varepsilon)} \right\rfloor = \text{closest integer to} \quad \frac{u}{\Delta(\varepsilon)}. \tag{6}$$

Note that $\kappa(0, \varepsilon) = 0$, and that (5) implies

$$\lim_{\varepsilon \to 0} \kappa(u, \varepsilon) = \infty \quad \text{and} \quad \lim_{\varepsilon \to 0} \frac{\kappa(u, \varepsilon)}{N(\varepsilon)} \to 0.$$

We estimate the mean $\mu$ of $X_t$ by the (observable) empirical mean

$$\bar{Y}^\varepsilon = \frac{1}{N} \sum_{n=1}^{N} Y_{n\Delta}^\varepsilon. \tag{7}$$

**Definition 3** For any given time lag $u$, let $\kappa = \kappa(u, \varepsilon)$ be as in Eq. (6). Define the time-shifted empirical mean $\tau_u \bar{Y}^\varepsilon$ by

$$\tau_u \bar{Y}^\varepsilon = \sum_{n=1}^{N} Y_{n\Delta}^\varepsilon.$$

We then estimate the lagged covariance matrix $K(u)$ of $X_t$ by the observable sub-sampled empirical covariances

$$\hat{K}_Y^\varepsilon(u) = \frac{1}{N} \left[ \sum_{n=1}^{N} Y_{n\Delta}^\varepsilon (Y_{(n+\kappa)\Delta}^\varepsilon)^* \right] - \bar{Y}^\varepsilon (\tau_u \bar{Y}^\varepsilon)^*, \tag{9}$$

where $N = N(\varepsilon), \Delta = \Delta(\varepsilon), \kappa = \kappa(u, \varepsilon)$

We always assume that $\varepsilon$ is small enough to force $N \gg \kappa$.

3.3 Sensitivity of Observable Covariances Estimators to Data Approximation

In formulas (7) and (9), replacing all observable $Y_t^\varepsilon$ by their limits $X_t$ transforms the observable estimators $\bar{Y}^\varepsilon, \tau_u \bar{Y}^\varepsilon$ and $\hat{K}_Y^\varepsilon(u)$ into unobservable estimators $\bar{X}^\varepsilon, \tau_u \bar{X}^\varepsilon$ and $\hat{K}_X^\varepsilon(u)$ given by

$$\bar{X}^\varepsilon = \frac{1}{N} \sum_{n=1}^{N} X_{n\Delta}, \tag{10}$$
\[ \tau_u \tilde{X}^e = \frac{1}{N} \sum_{n=1}^{N} X_{(n+\kappa)\Delta}, \]  
(11)

and

\[ \tilde{K}^e_X(u) = \frac{1}{N} \left[ \sum_{n=1}^{N} X_{n\Delta} X^e_{(n+\kappa)\Delta} \right] - \tilde{X}^e (\tau_u \tilde{X}^e)^* \]
\[ = \frac{1}{N} \sum_{n=1}^{N} (X_{n\Delta} - \tilde{X}^e) (X_{(n+\kappa)\Delta} - \tau_u \tilde{X}^e)^*, \]
(12)

where, as before, \( N = N(\varepsilon), \Delta = \Delta(\varepsilon), \kappa = \kappa(u, \varepsilon). \) Obviously, the covariance estimator remain unchanged when \( X_t \) is replaced by the centered process \( X_t - \mu. \) We now evaluate the \( L^2 \)-norm perturbations induced on \( \tilde{Y}^e \) and \( \tilde{K}^e_Y(u) \) when the observable data are replaced by their unobservable limits.

**Proposition 1** Consider any indirect observability setup \( X_t = \lim_{\varepsilon \to 0} Y_t^e \) as in Sect. 2. Let \( \rho(\varepsilon) \) be the \( L^4 \)-speed of convergence of \( Y_t^e \) to \( X_t. \) Let \( v \) be an upper bound for all the \( L^4 \) norms \( \| X_t \|_4 \) and \( \| Y_t^e \|_4. \) Fix an adaptive sub-sampling scheme \( \Delta(\varepsilon), N(\varepsilon) \) verifying (5). Then as \( \varepsilon \to 0, \) the observable estimators \( \tilde{Y}^e \) and \( \tilde{K}^e_Y(u) \) defined by (7) and (9) and their unobservable versions \( \bar{X}^e \) and \( \bar{K}^e_Y(u) \) verify

\[ \| \tilde{K}^e_Y(u) - \tilde{K}^e_X(u) \|_2 \leq 4 v \rho(\varepsilon) \text{ for all } u \geq 0, \]
\[ \| \tilde{Y}^e - \bar{X}^e \|_4 \leq \rho(\varepsilon) \text{ and } \| \tau_u \tilde{Y}^e - \tau_u \bar{X}^e \|_4 \leq \rho(\varepsilon). \]
(13) (14)

Proposition 1 reduces the \( L^2 \)-consistency analysis for our observable estimators of lagged moments to \( L^2 \)-consistency analysis for the unobservable moments estimators based on the \( X_t \) data and given by (10), (11), (12).

**Proof** We extend a similar proof given in [4]. The uniform bounds on \( \| Y_t^e - X_t \|_4, \| X_t \|_4, \| Y_t^e \|_4 \) are preserved by convex linear combinations, which yields

\[ \| \tilde{Y}^e \|_4 \leq v, \quad \| \bar{X}^e \|_4 \leq v, \quad \| \tilde{Y}^e - \bar{X}^e \|_4 \leq \rho(\varepsilon), \quad \| \tau_u \tilde{Y}^e - \tau_u \bar{X}^e \|_4 \leq \rho(\varepsilon). \]
(15)

Note that for any two random column vectors \( A, B \in \mathbb{R}^m, \) one has the elementary inequalities

\[ \| A B^* \|_2 \leq \| A \|_4 \| B \|_4 \quad \text{and} \quad \| A^* B \|_2 \leq m \| A \|_4 \| B \|_4 \]
(16)

Let \( V, W, V', W' \) be four random column vectors in \( \mathbb{R}^m, \) with \( L^4 \)-norms inferior to \( v, \) and verifying

\[ \| V - V' \|_4 \leq \rho \quad \text{and} \quad \| W - W' \|_4 \leq \rho. \]
(17)

By (16), the norms \( \| V (W - W')^* \|_2 \) and \( \| (V - V')(W')^* \|_2 \) are resp. inferior to \( \| V \|_4 \| W - W' \|_4 \) and \( \| W' \|_4 \| V - V' \|_4, \) so that

\[ \| V (W - W')^* \|_2 \leq v \rho \quad \text{and} \quad \| (V - V')(W')^* \|_2 \leq v \rho. \]

Hence, the random matrices \( V W^* \) and \( V'(W')^* \) verify

\[ \| V W^* - V'(W')^* \|_2 = \| V (W - W')^* + (V - V')(W')^* \|_2 \leq 2 v \rho. \]
(18)

The bound (18) can be applied when \( V, W, V', W' \) are replaced by \( \tilde{Y}^e, \tau_u \tilde{Y}^e, \bar{X}^e, \tau_u \bar{X}^e, \) and this yields

\[ \| \tilde{Y}^e (\tau_u \tilde{Y}^e)^* - \bar{X}^e (\tau_u \bar{X}^e)^* \|_2 \leq 2 v \rho(\varepsilon). \]
(19)
The bound (18) and relation (14) similarly show that
\[
\|Y_{n\Delta}^e (Y_{(n+\kappa)\Delta}^e)^* - X_{n\Delta} X_{(n+\kappa)\Delta}^*\|_2 \leq 2 \nu \rho (\varepsilon)
\]
(20)
for all \(n, \Delta, \kappa\). Convex combinations preserve this uniform \(L_2\)-bound, and hence
\[
\frac{1}{N} \sum_{n=1}^{N} \left\| Y_{n\Delta}^e (Y_{(n+\kappa)\Delta}^e)^* - X_{n\Delta} X_{(n+\kappa)\Delta}^* \right\|_2 \leq 2 \nu \rho (\varepsilon).
\]
(21)
By definitions of \(\hat{K}_T^e (u)\) and \(\hat{K}_X^e (u)\), the inequalities (19) and (21) conclude the proof. \(\square\)

4 \(L^2\)-consistency of Unobservable Sub-sampled Moments Estimators

Since Proposition 1 effectively controls in \(L^2\)-norm the difference between observable and unobservable estimators of lagged covariances, we now focus on the consistency of sub-sampled empirical covariance estimators based on the unobservable \(X_t\). This will require assuming a fast enough asymptotic decorrelation speed for the process \(X_t\).

4.1 Integrable Decorrelation Rates for 4th Order Moments

Definition 4 Let \(X_t \in \mathbb{R}^r\) be a stationary process with uniformly bounded moments of order \(4\). Denote by \(X_t (i), i = 1 \ldots r\) the coordinates of \(X_t\). For any time interval \(U \subset \mathbb{R}^+\), let \(F_U\) be the set of all random variables of the form \(X_s (i)\) or \(X_s (i)X_t (j)\) for arbitrary \(s, t \in U\) and \(1 \leq i, j \leq r\). Let \(f (T) > 0\) be a decreasing continuous function of \(T > 0\) with finite integral \(I (f) = \int_0^\infty f (T) \, dT\). We say that \(X_t\) has integrable decorrelation rate \(f (T)\) if for any disjoint time intervals \(U\) and \(V\) and for any random variables \(G \in F_U\) and \(H \in F_V\) one has
\[
\left| \mathbb{E} (GH) - \mathbb{E} (G) \mathbb{E} (H) \right| \leq f (T)
\]
(22)
where \(T = \min_{v \in V, u \in U} (v - u)\) is the time gap between \(U\) and \(V\).

Definition (22) controls up to moments of order \(4\) the “decorrelation” rate between \(X_t\) and \(X_{t+u}\) when \(u \to \infty\), and in particular implies that the lagged covariances matrices \(K (u)\) of the process \(X_t\) decay to zero at rate \(f (u)\) when \(u \to \infty\). The converse is of course not true for non-Gaussian processes \(X_t\), but when \(X_t\) is a Gaussian with lagged covariances \(K (u)\) decaying to zero at rate \(f (u)\), then \(X_t\) necessarily has integrable decorrelation rate proportional to \(f\), as can be seen from the classical formula
\[
\mathbb{E} (Z_1 Z_2 Z_3 Z_4) = \sigma_{1,2} \sigma_{3,4} + \sigma_{1,3} \sigma_{2,4} + \sigma_{2,3} \sigma_{1,4},
\]
where the four random variables \(Z_m\) are centered and jointly Gaussian with covariances \(\sigma_m, n\).

The decorrelation condition (22) is also linked with classical mixing contexts. Let \(\mathcal{F} (U)\) be the set of random variables measurable with respect to the sigma-algebra generated by all the \(X_s\) with \(s \in U\). To extend (22) to random variables \(G \in \mathcal{F} (U)\) and \(H \in \mathcal{F} (V)\) it is necessary to require a stronger “mixing” property for \(X_t\). Recall that when \(X_t\) is an ergodic stationary process having the \(\phi\)-mixing property (see, for instance, [16] for a recent survey), there is a fixed decay rate \(\phi (T) > 0\) tending to 0 as \(T \to \infty\), such that for any disjoint time intervals \(U\) and \(V\) separated by a time gap \(T > 0\), and for any pair of events \(A, B\) verifying \(1_A \in \mathcal{F} (U)\) and \(1_B \in \mathcal{F} (V)\) one has
\[
|P (B \mid A) - P (B)| \leq \phi (T).
\]
Provided the $X_t$ are in $L^4$, these uniform dependency decay rates will typically imply the validity of our condition (22) for some decorrelation function $f(T)$ deduced from $\phi(T)$.

4.2 Accuracy of Unobservable Sub-sampled Empirical Covariance Estimators

We now prove $L^2$- consistency for the unobservable sub-sampled empirical estimators of lagged covariance matrices. As could be expected from earlier results, integrable decorrelation rate for $X_t$ plays here a crucial role.

Proposition 2 Let $X_t \in \mathbb{R}^r$ be a stationary process with finite $L^4$- norm $\nu = \|X_t\|^4$, and with lagged covariance matrices $K(u)$ locally Lipschitz in $u$. Assume that $X_t$ has integrable decorrelation rate $f(T)$ as in (22) and set $I(f) = \int_0^{+\infty} f(T)dT$. Fix any adaptive sub-sampling scheme $\Delta(\varepsilon), N(\varepsilon)$ verifying (5). Let $\hat{K}_X^\varepsilon(u)$ be the (unobservable) sub-sampled empirical estimators of $K(u)$ defined by (12). Then the following two statements hold.

(I) For time lags $u$ in any bounded interval $[0, A]$, the estimators $\hat{K}_X^\varepsilon(u)$ are consistent in $L^2$ as $\varepsilon \to 0$, with uniform speed of convergence given by

$$\| \hat{K}_X^\varepsilon(u) - K(u) \|_2 \leq \gamma \left[ \frac{1}{\sqrt{N\Delta}} + \Delta \right],$$

where the constant $\gamma$ depends only on $A, I(f), \nu$.

(II) As $\varepsilon \to 0$, the (unobservable) estimators $\bar{X}^\varepsilon$ given by (10) converge in $L^4$ and in $L^2$ to the mean $\mu$ of $X_t$ with the speed of convergence given by

$$\| \bar{X}^\varepsilon - \mu \|_4 \leq \frac{c}{(N\Delta)^{1/4}} \quad \text{and} \quad \| \bar{X}^\varepsilon - \mu \|_2 \leq \frac{c}{(N\Delta)^{1/2}},$$

where the constant $c$ depends only on $I(f)$ and $\nu$.

Proof The proof relies on the meticulous use of natural techniques, and is hence given in the Appendix.

In Proposition 2, our key technical target was to identify how $L^2$ speeds of convergence depend on the adaptive scheme $N(\varepsilon), \Delta(\varepsilon)$, in order to optimize our adaptive subsampling schemes, first for our unobservable empirical estimators of lagged covariances (see next corollary), and further on for the corresponding observable covariances estimators.

4.2.1 Similarity Notation

Next, we introduce similarity notation for the limiting behavior of any two functions depending on a small parameter $\varepsilon$. In particular, for any two functions $G(\varepsilon)$ and $H(\varepsilon)$, we write $G \sim H$ whenever $\lim_{\varepsilon \to 0} \frac{G(\varepsilon)}{H(\varepsilon)}$ is finite and strictly positive.

The next proposition determines the optimal relationship between the number of observations $N(\varepsilon)$ and the sub-sampling time-step $\Delta(\varepsilon)$. Smaller values of $\Delta(\varepsilon)$ will not improve the limiting convergence rate, but will lead to oversampling and, thus, unnecessary data acquisition and storage.

Proposition 3 Let $X_t \in \mathbb{R}^r$ be a stationary process in $L^4$, with locally Lipschitz lagged covariances $K(u)$ and integrable decorrelation rate $f(T)$. Select arbitrary numbers of observations $N(\varepsilon)$ tending to $\infty$ as $\varepsilon \to 0$ and define the sub-sampling time steps by

$$\Delta(\varepsilon) \sim \frac{1}{N(\varepsilon)^{1/3}}.$$
The (unobservable) estimators $\hat{K}_X^\varepsilon(u)$ associated to $\Delta(\varepsilon)$, $N(\varepsilon)$ converge to the true $K(u)$ as $\varepsilon \to 0$, with the following $L^2$ speed, valid for all $0 \leq u \leq A$,
\[
\|\hat{K}_X^\varepsilon(u) - K(u)\|_2 \leq C \frac{1}{N(\varepsilon)^{1/3}},
\]
where $C$ is a constant determined by $A$, $I(f)$, $v$. Given the function $N(\varepsilon)$, the $L^2$ bounds in (26) define, up to multiplicative constants, the best $L^2$ speed of convergence obtainable under the generic assumptions of Proposition 2.

Proof For $\|\hat{K}_X^\varepsilon(u) - K(u)\|_2$, proposition 2 provides a bound proportional to
\[
B(\varepsilon) = \frac{1}{\sqrt{N\Delta}} + \Delta.
\]
For any given $N(\varepsilon)$, the choice $\Delta(\varepsilon) \sim N^{-1/3}(\varepsilon)$ obviously minimizes $B(\varepsilon)$, and one then has $B(\varepsilon) \sim N^{-1/3}(\varepsilon)$. This proves Eq. (26).

To check that the bound in Eq. (26) cannot be improved further, one simply needs to construct a process $X_t$ verifying all the hypotheses of Proposition 2, and such that there is an equivalence
\[
\|\hat{K}_X^\varepsilon(u) - K(u)\|_2 \sim \frac{1}{\sqrt{N\Delta}} + \Delta.
\]
Therefore, in this case the choice $\Delta \sim N^{-1/3}$ implies the optimal sub-sampling strategy. Indeed, in dimension 1, given any continuous and piecewise $C^1$ positive definite function $K(u)$, such that $|K(u)| \leq f(u)$ where $f(u)$ is continuous, decreasing, and integrable, there is a strictly stationary centered Gaussian process $X_t \in \mathbb{R}$ with covariances $K(u)$ (see [63]). The 2nd order moments of $X_t$ must then have integrable decorrelation rate proportional to $f$, as seen above. For standard examples of stationary 1-dimensional Gaussian processes, the estimators $\hat{K}_X^\varepsilon(u)$ do achieve $L^2$-errors of estimation which are actually equivalent to $(N\Delta)^{-1/2} + \Delta$. See for instance [6,7] where the case of the Ornstein-Uhlenbeck process is studied in detail.

5 Accuracy of Sub-sampled Moments Estimators Under Indirect Observability

We now study $L^2$-consistency for observable sub-sampled empirical estimators of lagged covariance matrices. The $L^4$-speed $\rho(\varepsilon)$ at which $Y_t^\varepsilon$ approximates $X_t$ is of course application specific. One of our goals was to analyze how $\rho(\varepsilon)$ determines optimal sub-sampling schemes for collecting observable data. This is achieved in the following result.

Theorem 1 Let $X_t = \lim_{\varepsilon \to 0} Y_t^\varepsilon$ be an indirect observability setup in $\mathbb{R}^r$ as in Sect. 2. Call $\mu$ and $K(u)$ the mean and lagged covariance matrices of $X_t$, respectively. Let $\rho(\varepsilon)$ be the $L^4$-speed at which $Y_t^\varepsilon$ approximates $X_t$ for all $t$. Let $v$ be an upper bound of all the $\|X_t\|_4$ and $\|Y_t^\varepsilon\|_4$. Assume that $X_t$ has integrable decorrelation rate $f(T)$.

For each adaptive sub-sampling scheme $\Delta(\varepsilon)$, $N(\varepsilon)$ verifying $\Delta(\varepsilon) \to 0$ and $N(\varepsilon) \Delta(\varepsilon) \to \infty$ as $\varepsilon \to 0$, formula (9) defines observable sub-sampled estimators $\hat{K}_X^\varepsilon(u)$ of $K(u)$. For each $u$, let $E(u)$ be the set of all such observable estimators.

Then among all the observable estimators $\hat{K}_X^\varepsilon(u)$ in $E(u)$, the best achievable $L^2$-speed of convergence to the true $K(u)$ as $\varepsilon \to 0$ is equivalent to some constant multiple of $\rho(\varepsilon)$ and is achieved by any sub-sampling scheme of the form.
\[
N(\epsilon) \sim \frac{1}{\rho(\epsilon)^3}, \quad \Delta(\epsilon) \sim \frac{1}{N(\epsilon)^{1/3}} \sim \rho(\epsilon).
\]  

For any such sub-sampling scheme and any fixed \(A\), one indeed has
\[
\|\hat{K}^e_y(u) - K(u)\|_2 \leq C\rho(\epsilon) \quad \text{for all } 0 \leq u \leq A,
\]
\[
\|\bar{Y}^e - \mu\|_2 \leq C\rho(\epsilon),
\]
for some constant \(C\) determined by \(A, v\) and \(I(f) = \int_1^{+\infty} f(T)dT\).

Let \(S(\epsilon) \sim N(\epsilon)\Delta(\epsilon)\) be the observational timespan gathering the time indexes \(n\Delta\) of all the observables \(Y^e_n\) involved in the estimator \(\hat{K}^e_y(u)\). Each optimized sub-sampling scheme of the form (27) also minimizes the rate at which \(S(\epsilon) \to \infty\) as \(\epsilon \to 0\).

Theorem 1 will be proved right after the following more technical proposition.

**Proposition 4** Let \(X_t = \lim_{\epsilon \to 0} Y^e_t\) be an indirect observability setup in \(\mathbb{R}^\nu\) verifying all the assumptions of Theorem 1. Let \(\rho(\epsilon)\) be the \(L^2\)-speed at which \(Y^e_t\) approximates \(X_t\) for all \(t\). Call \(\mu\) and \(K(u)\) the mean and lagged covariance matrices of \(X_t\). Moreover, fix any adaptive sub-sampling scheme \(\Delta(\epsilon), N(\epsilon)\) such that \(\lim_{\epsilon \to 0} N(\epsilon) = \infty\) and \(\Delta(\epsilon) \sim N^{-1/3}(\epsilon)\).

By formulas (9) and (7), this sub-sampling scheme defines observable estimators \(\hat{K}^e_y(u)\) and \(\bar{Y}^e\) of \(K(u)\) and \(\mu\), respectively. Then all these estimators are \(L^2\)-consistent as \(\epsilon \to 0\), and for any fixed \(A > 0\) one has the uniform \(L^2\)-speeds of convergence
\[
\|\hat{K}^e_y(u) - K(u)\|_2 \leq 4\rho(\epsilon) + \frac{c}{N^{2/3}(\epsilon)} \quad \text{for all } 0 \leq u \leq A,
\]
\[
\|\bar{Y}^e - \mu\|_2 \leq \rho(\epsilon) + \frac{c}{N^{2/3}(\epsilon)},
\]
where \(c\) is a constant determined by \(A, v, I(f)\).

**Proof** The bound in (30) is a direct consequence of the bounds obtained for \(\|\hat{K}^e_y(u) - \hat{K}^e_X(u)\|_2\) and \(\|\hat{K}^e_X(u) - K(u)\|_2\) in Propositions 1 and 2. Similar arguments prove the bound in (31).

We can now come back to proving Theorem 1.

**Proof of Theorem 1.** Taking \(\Delta \sim N^{-1/3}\), and hence \(S \sim N^{2/3}\), we apply Proposition 4. The bound \(B(\epsilon) = 4\rho + cN^{-1/3}\) in Eq. (30) is larger than \(4\rho\), so that to minimize \(B(\epsilon)\) up to multiplicative constants there is no asymptotic advantage in taking \(N^{-1/3} \ll \rho\). To simultaneously minimize (up to multiplicative constants) both \(S \sim N^{2/3}\) and \(B(\epsilon)\), a natural choice is then to take \(N^{-1/3} \sim \rho(\epsilon)\) which yields \(N \sim \rho^{-3}\) and hence \(\Delta \sim \rho\). Then, \(B(\epsilon)\) is inferior to \((4\rho + c)\rho\) which proves the Eq. (28). The bound on \(\|\bar{Y}^e - \mu\|_2\) provided by (31) is then equal to \((1 + c)\rho\) which proves the Eq. (29).

To show that the \(L^2\)-speeds of convergence in (28) and (29) cannot be generically improved for observable sub-sampled covariance matrix estimators in the class \(E(u)\), consider any 1-dimensional centered Gaussian process \(X_t\) with predefined covariance function \(K(u)\) assumed to be piecewise \(C^1\) and to decay at an integrable rate \(f(u)\) as \(u \to \infty\). Next, define \(Y^e_t = X_t + \rho(\epsilon)X_t\) where \(\rho(\epsilon)\) is any function such that \(\lim_{\epsilon \to 0} \rho(\epsilon) = 0\). Then \(\|Y^e_t - X_t\|_4 = C\rho(\epsilon)\) and all the hypotheses of Proposition 4 are satisfied. Moreover \(Y^e_t\) is a centered Gaussian process with lagged covariances \(K_X(u)(1 + 2\rho(\epsilon) + \rho^2(\epsilon))\). Then, for any adaptive sub-sampling scheme \(\Delta(\epsilon), N(\epsilon)\) the norm \(h(\epsilon)^2 = \mathbb{E}[(\hat{K}^e_y(u) - K(u))^2]\) can be explicitly computed in terms of \(N, \Delta, \rho\) and \(K(u)\), and as is easily checked, one always has \(\lim_{\epsilon \to 0} h(\epsilon)/\rho(\epsilon) > 0\). Since we already know that the optimized explicit sub-sampling
scheme (27) does yield $h(\varepsilon) \sim \rho(\varepsilon)$, this class of specific Gaussian examples proves generic optimality for the announced speed of convergence.

Previous results presented in this section have the following direct consequence for sparsely parametrized stationary processes.

**Theorem 2** Let $X_t = \lim_{\varepsilon \to 0} Y_t^\varepsilon$ be an indirect observability setup in $\mathbb{R}^r$ verifying all the assumptions of Theorem 1. Let $\rho(\varepsilon)$ be the $L^4$-speed at which $Y_t^\varepsilon$ approximates $X_t$ for all $t$. Assume also that $X_t$ is sparsely parametrized by $\theta \in \Theta \subset \mathbb{R}^p$, as in Definition 1. Then there exist observable estimators $\hat{\theta}^\varepsilon$ converging in probability to $\theta$ as $\varepsilon \to 0$. After selecting any sub-sampling scheme $N(\varepsilon), \Delta(\varepsilon)$ of the form (27), one can construct these estimators by an expression of the form

$$\hat{\theta}^\varepsilon = G(\hat{\Psi}^\varepsilon) \text{ with } \hat{\Psi}^\varepsilon = \left[ \hat{\Psi}_1^\varepsilon, \ldots, \hat{\Psi}_p^\varepsilon \right],$$

where $G : \mathbb{R}^p \to \mathbb{R}^p$ is a fixed smooth function and each $\hat{\Psi}_j^\varepsilon$ is an observable sub-sampled empirical estimator of the lagged 2nd order moment $\Psi_j$ of $X_t$. Moreover, if $\Theta$ is included in some known euclidean ball $\Lambda$ of finite radius, the truncated observable estimators $1_{\Lambda}(\hat{\theta}^\varepsilon)\hat{\theta}^\varepsilon$ then converge in $L^2$ norm to $\theta$ as $\varepsilon \to 0$, with $L^2$ speed of convergence faster than $C\rho(\varepsilon)$, for some constant $C$.

**Proof** Definition 1 states that $\theta = G(\Psi)$ where the vector $\Psi = \Psi_1, \ldots, \Psi_p$ involves $p$ lagged moments of order $\leq 2$ of the unobservable process $X_t$ and $G$ is a smooth function. Select an optimized adaptive sub-sampling scheme of the form (27), and let $\hat{\Psi}_j^\varepsilon$ be the associated observable sub-sampled empirical estimator of the lagged 2nd order moment $\Psi_j$ of $X_t$. By Theorem 1, as $\varepsilon \to 0$, the estimators $\hat{\Psi}_j^\varepsilon$ converge to $\Psi_j$ in $L^2$, and hence also converge to $\Psi_j$ in probability. Since convergence in probability is preserved by smooth functions, the estimators $\hat{\theta}^\varepsilon$ by Eq. (32) must then converge in probability to $\theta$ as $\varepsilon \to 0$. For uniformly bounded random vectors, convergence in probability always implies convergence in $L^2$, and this proves the last statement of the theorem.

**6 Applications to Indirectly Observable Multi-dimensional Diffusions**

We now present several examples of stationary multi-dimensional diffusions $X_t$ naturally embedded in indirect observability frameworks. A key assumption to prove $L^2$-consistency for natural observable estimators of the lagged covariances $K(u)$ of $X_t$ is to require the lagged moments of order $\leq 4$ of $X_t$ to decay at integrable decorrelation rate for large lags, as specified in Eq. (22).

Published literature does not provide easy generic conditions on SDEs coefficients guaranteeing that the associated diffusion $X_t$ has integrable decorrelation rate as specified in (22). Quite relevant exponential decay bounds for the transition density $p_\theta(t, x, y)$ as $t \to \infty$ have been given in [3, 23, 52, 60], but more precise bounds on $p_\theta$ are needed to generically validate the integrable decorrelation rates on lagged moments of order 4 as required by Eq. (22). Here we do not attempt to solve these technical questions for general classes of diffusions. Instead, we will simply list a few interesting examples of diffusions for which our assumptions can either be directly verified, or are quite plausibly conjectured to be true, as can be also tested by numerical simulations.
6.1 Gradient Diffusions

In [36, Section 7.1], M. Hairer discusses the “gradient diffusions” \( X_t \) in \( \mathbb{R}^r \) driven by the SDE

\[
dX_t = -\nabla Q(X_t) dt + \sigma dW_t,\]

where \( Q(x) \) is a smooth “potential” defined for \( x \in \mathbb{R}^r \), and \( \sigma \) is a constant \( r \times r \) matrix. The potential \( Q \) is also assumed to behave as a polynomial at infinity, i.e. there are constants \( c, C, k \) such that

\[
c|x|^{2k} \leq Q(x) \leq C|x|^{2k}, \quad \langle x, \nabla Q(x) \rangle \geq c|x|^{2k}, \quad |D^2Q(x)| \leq C|x|^{2k-2},
\]

where \( \nabla \) is the gradient operator and \( D \) denotes 1st order differentiation operators. Under these conditions, [36] proves that the probability distribution of \( X_t \) given \( X_0 = x \) converges to the stationary probability distribution of \( X_t \) at exponentially fast speed as \( t \to \infty \), and that \( X_t \) verifies the classical Doeblin property. Results of [36] imply the exponentially fast decorrelation

\[
|E(GH) - E(G)E(H)| \leq e^{-\gamma T},
\]

where \( \gamma > 0 \) is a constant, but only for random variables \( G, H \) of the form \( G = f(X_s)\phi(X_t) \) and \( H = F(X_u)\Psi(X_v) \) with \( f, \phi, F, \Psi \) bounded and \( s \leq t < t + T \leq u \leq v \). Whenever the Aronson bounds [3] on transition densities hold, this decorrelation inequality can be extended to \( G = X_sX_t \) and \( H = X_uX_v \) so that the “gradient diffusions” \( X_t \) provide a class of stationary diffusions with integrable decorrelation rate \( f(T) = e^{-\gamma T} \) and finite moments of order 4. Whenever a given multidimensional diffusion process \( X_t \) verifies all our assumptions on the unobservable process, the simplest example of observable processes \( Y_{t}^{\varepsilon} \) associated to \( X_t \) is generated by local smoothing, i.e.

\[
Y_{t}^{\varepsilon} = \frac{1}{\varepsilon} \int_{t-\varepsilon}^{t} X_s ds.
\]

Such observable processes are often generated by sensors recording short-term averages of high frequency input data. See [6] for a detailed study of this case when \( X_t \) is a Gaussian diffusion.

6.2 Volatility Processes and Heston Joint SDEs

A striking example of indirect observability is quite ubiquitous in stochastic modeling of joint price and volatility for stockmarket data. The well known Heston model (see [38]) links the price \( S_t \) and the squared volatility \( V_t \) of an asset by parametrized joint SDEs of the form

\[
dS_t = \mu S_t dt + \sqrt{V_t} S_t dW_1,
\]

\[
dV_t = \kappa(\theta - V_t) dt + \sigma \sqrt{V_t} dW_2,
\]

where the unknown positive parameters \( \mu, \kappa, \theta, \sigma \) need to be estimated from asset price data \( S_t \) only, since the squared instantaneous volatility \( V_t \) is not directly available, and plays the part of our unobservable process \( X_t = V_t \). In our indirect observability framework, volatility approximations \( Y_{t}^{\varepsilon} \) based either on prices \( S_t \) or on observed option prices become

\[
\text{(c) Springer}
\]
the observable processes. A classical volatility approximation is the “realized volatility” given by the sum of squared returns

\[ Y^\varepsilon_t = \frac{1}{M \varepsilon} \sum_{k=1}^{M} (R_{tk} - R_{tk-1})^2, \quad \text{where} \quad dR_t = \frac{dS_t}{S_t}, \quad (33) \]

In this equation, the time step size \( \varepsilon \) and the window size \( M = M(\varepsilon) \) are user selected. The \( L^2 \) convergence of realized volatility to instantaneous volatility as \( \varepsilon \to 0 \) is studied in [12,13]. In a companion paper to be published in [5], we have proved that the pair \( (Y^\varepsilon_t, V_t) \) verifies all the hypotheses of our indirect observability setup, and we have completed a detailed analysis of parameter estimation under indirect observability for generic Heston models (see also [62]).

6.3 Averaged Multiscale Stochastic Systems

Consider a “slow-fast” joint SDEs system (see [57] for overview and references) involving a (small) scale parameter \( \varepsilon \) and given by

\[
\begin{align*}
    dx_t &= a(x_t, y_t)dt + b(x_t, y_t)dW_1(t), \\
    dy_t &= \frac{1}{\varepsilon}c(x_t, y_t)dt + \frac{1}{\sqrt{\varepsilon}}d(x_t, y_t)dW_2(t),
\end{align*}
\]

where \( W_1(t) \) and \( W_2(t) \) are independent Brownian motions and the coefficients \( a, b, c, d \) are bounded smooth functions of \( x, y \). Note that the diffusions \( x_t, y_t \) actually depend on the scale parameter \( \varepsilon \). Assume that for any fixed \( x \), the “fast” SDE driving \( y_t \) has a stationary distribution \( q(y|x) \) verifying \( \mathbb{E}_q a(x, y) \neq 0 \). Then under mild complementary conditions on \( a, b, c, d \), and as \( \varepsilon \to 0 \), the process \( x_t \) converges in probability to the “reduced dynamics”

\[
    dX_t = A(X_t)dt + B(X_t)BB^*(X_t)dW_1(t),
\]

where \( A = \mathbb{E}_q a(x, y) \) and \( BB^* = \mathbb{E}_q bb^* \). Convergence in probability implies \( L^4 \) convergence for variables bounded in \( L^4 \), and the \( L^3 \) convergence of \( x_t \) to \( X_t \) is proved in [57] for periodic coefficients and \( x_t, y_t \) on a torus. In practical applications, one essentially wants to parametrize the slow asymptotic SDE (36) driving the unobservable process \( X_t \), and the only realistically accessible data are generated by the approximating process \( x_t \), since for small \( \varepsilon \) the \( y_t \) data are too noisy to be reliably acquired. Hence the slow process \( x_t \) plays the role of the observable process \( Y^\varepsilon_t \).

There are many practical applications when \( A(X_t) \) has polynomial nonlinearities. When \( X_t \) is one-dimensional this trivially corresponds to the case of gradient diffusions discussed earlier in this section. For multi-dimensional \( X_t \) one can conjecture that the mixing rates for the process \( X_t \) must obey exponential decay unless Eq. (36) posses some unusual properties (e.g. special symmetries or existence of conserved quantities). It is then quite reasonable to expect exponential convergence to the equilibrium distribution and, thus, exponentially fast decorrelation rates of lagged 4th moments for large lags. Moreover, exponentially fast decorrelation rates have been demonstrated numerically in many practical examples. So we expect that our key assumptions on the unobservable and observable processes will be satisfied for many multiscale examples of the form (36) and (34), respectively.
7 Conclusions

For stationary processes $X_t \in \mathbb{R}^r$ which are not directly observable, but can be approximated in $L^4$ by observable processes $Y^\varepsilon_t$ as $\varepsilon \to 0$, we have developed a mathematical framework where the unknown vector $\theta \in \mathbb{R}^P$ parametrizing $X_t$ can be consistently and efficiently estimated from adequately sub-sampled observations of $Y^\varepsilon_t$. The present paper extends substantially several of our earlier results [4,6,7] to non-Gaussian stationary processes $X_t$.

We have focused on the ‘‘sparsely parametrized’’ situations where $\theta \in \mathbb{R}^P$ is a (generally non explicit) smooth function $G(\Psi_1, \ldots, \Psi_P)$ of $p$ lagged moments of order $\leq 2$ of $X_t$. We conjecture that this holds true when $X_t$ is a stationary multi-dimensional diffusion provided the matrix diffusion coefficients $a(x, \theta)$ and the drift $b(x, \theta)$ of $X_t$ are analytic in $x$ and $\theta$.

The above setup leads us to study the class of parameter estimators of the form $\hat{\theta}^\varepsilon = G(\hat{\Psi}^\varepsilon)$, where $\hat{\Psi}^\varepsilon$ is an observable empirical estimator of $\Psi$ based on the $N(\varepsilon)$ sub-sampled observable data $Y^\varepsilon_{n\Delta(\varepsilon)}$, with $n = 1 \ldots N(\varepsilon)$. For parameter estimators such as $\hat{\theta}^\varepsilon$, analysis of consistency and speed of convergence is essentially equivalent to a similar but more technical analysis for observable subsampled estimators $\hat{K}_{Y^\varepsilon(u)}$ of the lagged covariances $K(u)$ of $X_t$. Note that for $u > 0$, estimators $\hat{K}_{Y^\varepsilon(u)}$ involve only non-vanishing time lags $u(\varepsilon)$ (with $u(\varepsilon) \to u > 0$ as $\varepsilon \to 0$), since vanishing time lags decrease robustness to data perturbations (see [4,6]).

We explicitly determine how to best choose the sub-sampling time step $\Delta(\varepsilon)$ and the number $N(\varepsilon)$ of observations in terms of the $L^4$ distance $\rho(\varepsilon) = ||Y^\varepsilon_t - X_t||_4$ (see Eq. (27)). Our asymptotically optimal sub-sampling schemes $N(\varepsilon) \sim \rho^{-3}(\varepsilon)$ and $\Delta(\varepsilon) \sim \rho(\varepsilon)$ are constructed to simultaneously minimize the amplitude of estimation errors as well as the computational/observational complexity due to both the number $N(\varepsilon)$ and the time span $S(\varepsilon) = N(\varepsilon)\Delta(\varepsilon)$ of sub-sampled observable data. Indeed, in many practical situations such as joint dynamic modeling of observable stock prices and unobservable volatilities, both $N(\varepsilon)$ and $S(\varepsilon)$ must remain rather moderate, even for intraday data. Our optimal sub-sampling results rely on a key hypothesis, stating that for $s \leq t \leq u \leq v$ the random variables $X_s X_t$ and $X_{u+T} X_{v+T}$ decorrelate at an integrable rate $f(T) \to 0$ when $T \to \infty$ (see (22)). This is generally true for the many practical situations where $X_t$ is a stationary multi-dimensional diffusion with exponentially fast mixing.

When $X_t$ is sparsely parametrized and has integrable decorrelation rate $f$, we prove that as $\varepsilon \to 0$, the sub-sampled observable estimators of lagged covariances determined by our optimal sub-sampling scheme (27) converge in $L^2$ to the true lagged covariances of $X_t$, with $L^2$-speeds of convergence faster than $c\rho(\varepsilon)$ for some constant $c$, and that the decorrelation rate $f$ only affects the constant $c$ via the integral $I(f)$ of $f$. Our associated observable subsampled parameter estimators $\hat{\theta}^\varepsilon = G(\hat{\Psi}^\varepsilon)$ are then consistent in probability when $\varepsilon \to 0$. In practical applications, the unknown $\theta$ is of course a priori bounded, and then a natural truncation of the estimators $\hat{\theta}^\varepsilon$ guarantees their $L^2$-convergence to $\theta$ at $L^2$-speed faster than some constant multiple of $\rho(\varepsilon)$.

Our work thus points out the pragmatic impact of numerical methods enabling fast evaluation of $\rho(\varepsilon)$, to help determine nearly optimal sub-sampling schemes, as well as for computing approximate error bars on parameter estimators. We will study numerical applications of our approach for non-Gaussian $X_t$ in subsequent papers.

Our indirect observability study has strong practical consequences for a broad range of applications. Let us mention two examples. In financial mathematics, our indirect observability setup potentially applies to many stochastic volatility models driving the price and volatility of a given asset. The observable $Y^\varepsilon_t$ can then be a realized volatility estimated on
a time window depending on \( \varepsilon \), and \( X_t \) is the unobservable instantaneous volatility. For the well known Heston joint SDEs, our approach has enabled the construction of consistent and efficient explicit parameter estimators based on optimally sub-sampled realized volatility data [5, 6].

A second class of examples concerns complex multiscale systems driving atmospheric or ocean dynamics. In this case the numerical datasets generated by known high dimensional fluid evolution models can be analyzed by artificially inserting a small scaling parameter \( \varepsilon \) into the model to further accelerate the fast variables and numerically analyze (as \( \varepsilon \) varies) the behavior of parameter estimators for key parameters of the slow dynamics. This computer intensive version of our approach should yield both a concrete and efficient optimal sub-sampling scheme as well as approximate error bars for our parameter estimators. We will present detailed actual examples in further publications.

Acknowledgments I.T. and R.A. were supported in part by the NSF Grant DMS-1109582.

Appendix: \( L^2 \)-consistency for Unobservable Estimators of Lagged 2nd Order Moments

In this appendix we present a detailed Proof of Theorem 2, which addresses the \( L^2 \)-consistency results for the unobservable sub-sampled empirical estimators \( \bar{X}^\varepsilon \) and \( \hat{K}_X^\varepsilon(u) \) of means and lagged covariances. The hypotheses and notations are those of Theorem 2. Replacing \( X_t \) by the centered process \( X_t - \mu \) and setting \( \mu = 0 \) is a trivial change in the proof, so we only need to consider the case where all \( X_t \) are centered and \( \mu = 0 \).

Step 1 Sums of decorrelation values For all \( D > 0 \) and \( j \geq 1 \) one has \( Df(jD) \leq \int_{(j-1)D}^{jD} f(T)dT = I(f)/D \). This implies

\[
\sum_{j=1}^{\infty} f(jD) \leq \frac{1}{D} \sum_{j=1}^{\infty} \int_{(j-1)D}^{jD} f(T)dT = I(f)/D. \tag{37}
\]

Define the function \( g(q, D) \) for all integers \( q \geq 2 \) and all \( D > 0 \) by

\[
g(q, D) = \sum_{1 < m < n \leq 1+q} f((n-m)D) = \sum_{j=1}^{q-1} jf(jD). \tag{38}
\]

Due to (37), the following inequality holds for all \( D > 0 \) and \( q \geq 2 \)

\[
g(q, D) \leq (q-1) \sum_{j=1}^{q-1} f(jD) \leq (q-1)I(f)/D. \tag{39}
\]

Step 2 Sub-sampled empirical means converge in \( L^2 \) Fix an integer \( j \in [1 \ldots r] \). Denote the \( j \)-th coordinates of \( X_{n\Delta} \) and of the empirical mean estimator \( \bar{X}^\varepsilon \) by

\[
U_n = X_{n\Delta}(j) \quad \text{and} \quad \bar{X}^\varepsilon(j) = \frac{1}{N}(U_1 + \cdots + U_N).
\]

With the notation \( s_j^2 = \mathbb{E}(U_n^2) \), this implies

\[
N^2\mathbb{E}[(\bar{X}^\varepsilon(j))^2] = Ns_j^2 + 2 \sum_{1 \leq m < n \leq N} \mathbb{E}[U_mU_n]. \tag{40}
\]
Applying the decorrelation hypothesis (22) and the relations (38), (39), we obtain

\[ \left| \sum_{1 \leq m < n \leq N} E(U_m U_n) \right| \leq \sum_{1 \leq m < n \leq N} f((n - m)\Delta) = g(N - 1, \Delta) \leq I(f) \frac{N}{\Delta}. \]

The definition of the \( L^q \)-norm also implies \( s_j \leq \|X_t\|_2 \leq \|X_t\|_4 = v \). Hence, (40) implies

\[ (\|X^e(j)\|^2)^2 \leq \frac{v^2}{N} + \frac{2I(f)}{N\Delta}. \]

For any \((r_1 \times r_2)\) random matrix \(M\), and any \(q \geq 1\) our definition of the norm \(\|M\|^q\) implies

\[ \|M\|_q \leq (r_1 r_2)^{1/q} \max_{i,j} \|M_{i,j}\|_q. \]  

(41)

The inequality \( \|X^e\|_2 \leq \sqrt{r} \max_j \|X^e(j)\|_2 \), then yields, due to (41),

\[ \|X^e\|_2 \leq \sqrt{r} \left( \frac{v^2}{N} + \frac{2I(f)}{N\Delta} \right)^{1/2} \leq \frac{1}{\sqrt{N\Delta}} (v(r\Delta)^{1/2} + (2rI(f))^{1/2}) \cdot \]

Since \( \Delta(\varepsilon) \to 0 \) this proves the \( L^2 \)-bound in (24) when \(X_t\) is centered and hence in general.

**Step 3 Sub-sampled empirical means converge in \( L^4 \)** Basic algebra yields the identities

\[ N^4 (\bar{X}^e(j))^4 = \sum_{m,n,a,b \in [1...N]} U_a U_b U_m U_n = S_1 + S_2 + 2S_3 + 24S_4, \]  

(42)

where the sums \(S_1, S_2, S_3,\) and \(S_4\) are defined by

\[ S_1 = \sum_{1 \leq m \leq N} U^4_m, \]

\[ S_2 = \sum_{1 \leq m < n \leq N} \left[ 2U^2_m U^2_n + U^3_m U_n + U^2_m U^2_n \right], \]

\[ S_3 = \sum_{1 \leq a < m < n \leq N} \left[ U^2_a U_m U_n + U_a U^2_m U_n + U_a U_m U^2_n \right], \]

\[ S_4 = \sum_{1 \leq a < b < m < n \leq N} U_a U_b U_m U_n. \]

Due to the assumption that the \( L^4 \) norm of \(X_t\) is bounded uniformly by \(v\), one clearly has \(|E(S_1)| \leq Nv^4\) and \(|E(S_2)| \leq 4N^2v^4\).

Since we are considering the centered process \(X_t, E(U_n) \equiv 0\), and for \(a < m < n\) the decorrelation hypothesis implies

\[ |E(U^2_a U_m U_n)| = |E(U^2_a U_m U_n) - E(U^2_a U_m)E(U_n)| \leq f((n - m)\Delta). \]

Similarly, one shows that

\[ |E(U_a U^2_m U_n)| \leq f((n - m)\Delta) \quad \text{and} \quad |E(U_a U^2_m U_n)| \leq f((m - a)\Delta). \]

These bounds and definition (38) yield (for \(N \geq 3\))

\[ |E(S_3)| \leq \sum_{1 \leq a < m < n \leq N} [2f((n - m)\Delta) + f((m - a)\Delta)] = 3 \sum_{2 \leq m \leq N-1} g(m, \Delta) \]
which implies, due to the bound (39),

$$|E(S_3)| \leq 3I(f) \sum_{2 \leq m \leq N-1} (m - 1) \leq \frac{3I(f)N^2}{2\Delta}.$$  

As above, one also has $|E(U_a U_b U_m U_n)| \leq f((b - a)\Delta)$ for $a < b < m < n$. The expressions of $S_4$ and $g(m, \Delta)$ then yield (for $N \geq 4$)

$$|E(S_4)| \leq \sum_{1 \leq a < b < m < n \leq N} f((b - a)\Delta) = \sum_{3 \leq m < n \leq N} g(m, \Delta) = \sum_{3 \leq m \leq N-1} (N - m)g(m, \Delta).$$

Therefore, due to (39) we obtain for $N \geq 4$

$$|E(S_4)| \leq \sum_{3 \leq m \leq N-1} (N - m)I(f) \frac{m}{\Delta} \leq \frac{I(f)N^3}{4\Delta}.$$  

Finally, the bounds on $|E(S_k)|$, and Eq. (42) entail

$$E\left(\left(\bar{X}^\varepsilon(j)\right)^4\right) \leq \frac{5\nu^4}{N^2} + \frac{3I(f)N^2}{N\Delta} + \frac{6I(f)N}{N\Delta} \leq \frac{C}{N\Delta}$$

(43)

for some explicit constant $C$, since $N(\varepsilon) \to \infty$ and $\Delta(\varepsilon) \to 0$ with $N(\varepsilon)\Delta(\varepsilon) \to \infty$. In particular for $\varepsilon$ small enough, one can clearly take $C = 7I(f)$. Therefore, Eqs. (41) and (43) imply

$$\|\bar{X}^\varepsilon\|_4 \leq r^4 \max_{i,j} \|\bar{X}^\varepsilon\|_4 \leq \frac{(rC)^{1/4}}{(N\Delta)^{1/4}}$$

which proves the $L^4$-bound in (24).

**Step 4 Convergence of empirical lagged covariance matrices estimators** Introduce the short-hand notations $V_n = X_{n\Delta}$ and

$$\bar{V}_N = \bar{X}^\varepsilon = \frac{1}{N} \sum_{n=1}^N V_n,$$  

(44)

$$\tau \bar{V}_N = \tau \bar{X}^\varepsilon = \frac{1}{N} \sum_{n=1}^N V_{n+\kappa},$$  

(45)

$$W_N = \frac{1}{N} \sum_{n=1}^N V_n V_{n+\kappa}^*.$$  

(46)

From the Definition (12), the covariance matrix estimators $\hat{K}_X^\varepsilon(u)$ can be rewritten as

$$\hat{K}_X^\varepsilon(u) = W_N - \bar{V}_N(\tau \bar{V}_N)^*.$$  

(47)

First, we evaluate the term $\bar{V}_N(\tau \bar{V}_N)^*$ in the equation above. Impose $0 \leq u \leq A$ for some fixed $A$. Thus, by construction

$$\|\bar{V}_N - \tau \bar{V}_N\|_4 \leq 2\kappa \nu/N \leq 2(u + \Delta)\nu/N \leq 2(1 + A)\nu \frac{1}{N\Delta}$$

and applying the inequality (18) one arrives at the following relation

$$\|\bar{V}_N(\tau \bar{V}_N)^* - \bar{V}_N(\bar{V}_N)^*\|_2 \leq 4(1 + A)\nu^2 \frac{1}{N\Delta}.$$  

(48)
Since \( \mu = 0 \), we also have
\[
\| \tilde{V}_N \|_4 \leq \frac{C}{(N \Delta)^{1/4}},
\]
as proven in Step 3. This implies, by inequality (18),
\[
\| \tilde{V}_N (\tilde{V}_N)^* \|_2 \leq 2 \left[ \frac{C}{(N \Delta)^{1/4}} \right]^2 = \frac{2C^2}{(N \Delta)^{1/2}}
\]
which yields, due to Eq. (48),
\[
\| \tilde{V}_N (\tau \tilde{V}_N)^* \|_2 \leq \frac{2C^2}{\sqrt{N \Delta}} + \frac{4(1 + A)v^2}{N \Delta}. \tag{49}
\]
By the construction of \( \kappa(u, \varepsilon) \), the “discrete” lag \( \kappa \Delta \) is close to continuous lag \( u \) and \( |\kappa \Delta - u| \leq \Delta \). Since the true lagged covariance matrices \( K(u) \) are locally Lipschitz, there is a constant \( \lambda = \lambda(A) \) such that for all \( 0 \leq u \leq A \) and all \( \varepsilon > 0 \) the following deterministic inequality holds
\[
\| K(u) - K(\kappa \Delta) \| \leq \lambda |u - \kappa \Delta| \leq \lambda \Delta. \tag{50}
\]
Next, we compare the term \( W_N \) in the expression of the covariance estimator (47), with the true covariance matrix \( K(\kappa \Delta) \) evaluated at the “discretized” time lag \( \kappa \Delta \). Since \( X_t \) is stationary, we have \( K(\kappa \Delta) = \mathbb{E}(V_n V_{n+k}^*) \) for all \( n \), and formula (46) implies that
\[
W_N - K(\kappa \Delta) = \frac{1}{N} \sum_{n=1}^{N} \left( V_n V_{n+k}^* - \mathbb{E}[V_n V_{n+k}^*] \right).
\]
For any two coordinates \( i, j \in \{1, \ldots, r\} \) denote \( T_n = V_n(i) \) and \( U_n = V_n(j) \) as the \( i \)-th and \( j \)-th coordinates of \( V_n \), respectively. In addition, we also define
\[
H_n = T_n U_{n+k} - \mathbb{E}[T_n U_{n+k}].
\]
Clearly \( \mathbb{E}[H_n] = 0 \) and the \( (i, j) \) coefficient of the matrix \( M = W_N - K(\kappa \Delta) \) is then
\[
M_{i, j} = \frac{1}{N} \sum_{n=1}^{N} H_n
\]
and
\[
N^2 \mathbb{E}[M_{i, j}^2] = \sum_{(m, n) \in [1 \ldots N]} \mathbb{E}[H_m H_n]. \tag{51}
\]
Next, we partition the summation interval in the expression above into two complementary sets, \( Q^+ \) and \( Q^- \), as follows
\begin{itemize}
  \item \( (m, n) \in Q^+ \) whenever \( |n - m| > \kappa \) and \( m, n \in [1 \ldots N] \),
  \item \( (m, n) \in Q^- \) whenever \( |n - m| \leq \kappa \) and \( m, n \in [1 \ldots N] \).
\end{itemize}
Due to bounded fourth moments of the process \( X_t \), we have \( |\mathbb{E}[H_m H_n]| \leq 2v^4 \). Moreover, cardinal (\( Q^- \)) = \( N + \kappa (2N - \kappa - 1) \) \( \leq 3N \kappa \), and, therefore,
\[
\left| \sum_{(m, n) \in Q^-} \mathbb{E}(H_m H_n) \right| \leq 6v^4 N \kappa. \tag{52}
\]
For \((m, n) \in Q^+\), the decorrelation rate of the 2nd order moments yields

\[|E[H_m H_n]| \leq f(|n - m|\Delta),\]

so that

\[\left| \sum_{(m,n) \in Q^+} E[H_m H_n] \right| \leq \sum_{(m,n) \in Q^+} f(|n - m|\Delta).\] (53)

Thus, relation (51) and inequalities (52), (53) imply

\[N^2 E[M_{i,j}^2] \leq \sum_{(m,n) \in Q^+} f(|n - m|\Delta) + 6\nu^4 N\kappa.\] (54)

Easy algebra transforms the double sum above into

\[\sum_{(m,n) \in Q^+} f(|n - m|\Delta) = 2 \sum_{\kappa+1 \leq s \leq N-1} (N-s)f(s\Delta) \leq 2N \sum_{1 \leq s \leq N-1} f(s\Delta) \leq 2I(f)\frac{N}{\Delta},\]

where Eq. (37) were used in the last inequality. Recall that for \(0 \leq u \leq A\) and due to the construction of \(\kappa\), one also has

\[\kappa \leq \frac{u}{\Delta} + \Delta \leq \frac{A+1}{\Delta}.\]

Substituting the last two expressions into (54) we obtain

\[E[M_{i,j}^2] \leq (2I(f) + 6(A + 1)\nu^4) \frac{1}{N\Delta}.\]

By Eq. (41) we further obtain

\[\|W_N - K(\kappa\Delta)\|_2 = \|M\|_2 \leq r \left(2I(f) + 6(A + 1)\nu^4\right)^{1/2} \frac{1}{\sqrt{N\Delta}}.\] (55)

Using the expression for \(\hat{K}_{\chi}^\varepsilon(u)\) in (47) and the triangle inequality we can write

\[\|\hat{K}_{\chi}^\varepsilon(u) - K(u)\|_2 \leq \|W_N - K(\kappa\Delta)\|_2 + \|K(\kappa\Delta) - K(u)\|_2 + \|\hat{V}_N(\tau \hat{V}_N)^*\|_2.\]

Combining the three bounds in (49), (50), and (55), we obtain, for all \(\varepsilon > 0\) and \(0 \leq u \leq A\),

\[\|\hat{K}_{\chi}^\varepsilon(u) - K(u)\|_2 \leq \frac{\Gamma}{\sqrt{N\Delta}} + \lambda\Delta,\] (56)

where

\[\Gamma = 2C^2 + r \left(2I(f) + 6(A + 1)\nu^4\right)^{1/2} + \frac{4(1 + A)\nu^2}{\sqrt{N\Delta}}.\]

Moreover, for \(\varepsilon\) small enough, we can take \(C^2 = \sqrt{\tau I(f)}\) as discussed in Step 2, and \(4(1 + A)\nu^2(N\Delta)^{-1/2}\) will become much smaller than \(\sqrt{\tau I(f)}\). Therefore, for \(\varepsilon\) small enough, one has (using \(\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}\)) a simplified expression for the constant \(\Gamma\)

\[\Gamma \leq \gamma = 8\sqrt{\tau I(f)} + 2.5\nu^2\sqrt{A + 1}.\]

This concludes the Proof of Theorem 2.
References

1. Aït-Sahalia, Y.: Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica 70, 223–262 (2002)
2. Aït-Sahalia, Y., Mykland, P.A., Zhang, L.: How often to sample a continuous-time process in the presence of market microstructure noise. Rev. Financ. Stud. 18(2), 351–416 (2005)
3. Aronson, D.G.: Bounds for the fundamental solution of a parabolic equation. Bull. Am. Math. Soc. 73(6), 890–896 (1967)
4. Azencott, R., Beri, A., Jain, A., Timofeyev, I.: Sub-sampling and parametric estimation for multiscale dynamics. Commun. Math. Sci. 11(4), 939–970 (2013)
5. Azencott, R., Beri, A., Ren, P., Timofeyev, I.: Parametric estimation of the volatility equation in the Heston model using indirect observations. In Preparation (2015)
6. Azencott, R., Beri, A., Timofeyev, I.: Adaptive sub-sampling for parametric estimation of Gaussian diffusions. J. Stat. Phys. 139(6), 1066–1089 (2010)
7. Azencott, R., Beri, A., Timofeyev, I.: Parametric estimation of stationary stochastic processes under indirect observability. J. Stat. Phys. 144(1), 150–170 (2011)
8. Azencott, R., Dacunha-Castelle, D.: Series of Irregular Observations: Forecasting and Model Building. Springer, New York (1986)
9. Azencott, R., Gadhyan, Y.: Accuracy of maximum likelihood parameter estimators for Heston stochastic volatility sde. J. Stat. Phys. 159(2), 393–420 (2015)
10. Bandi, F.M., Phillips, P.C.B.: Fully nonparametric estimation of scalar diffusion models. Econometrica 71(1), 241–283 (2003)
11. Barndorff-Nielson, O.E., Hansen, P.R., Lunde, A., Shephard, N.: Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. Econometrica 76(6), 1481–1536 (2008)
12. Barndorff-Nielson, O.E., Shephard, N.: Econometric analysis of realized volatility and its use in estimating stochastic volatility models. J. R. Stat. Soc. 64, 253–280 (2002)
13. Barndorff-Nielson, O.E., Shephard, N.: Power and bipower variation with stochastic volatility and jumps. J. Financ. Econ. 2(1), 1–37 (2004)
14. Berner, J.: Linking nonlinearity and non-Gaussianity of planetary wave behavior by the Fokker-Planck equation. J. Atmos. Sci. 62, 2098–2117 (2005)
15. Bollen, B., Inder, B.: Estimating daily volatility in financial markets utilizing intraday data. J. Empir. Financ. 9(5), 551–562 (2002)
16. Bradley, R.C.: Basic properties of strong mixing conditions. A survey and some open questions. Probab. Surv. 2, 107–144 (2005)
17. Brandt, M.W., Santa-Clara, P.: Simulated likelihood estimation of diffusions with an application to exchange rate dynamics in incomplete markets. J. Financ. Econ. 63, 161–210 (2002)
18. Christensen, K., Podolskij, M.: Asymptotic theory of range-based multipower variation. J. Financ. Economet. (2012)
19. Comte, F., Genon-Catalot, V., Rozenholc, Y.: Penalized nonparametric mean square estimation of the coefficients of diffusion processes. Bernoulli 13(2), 514–543 (2007)
20. Crommelin, D., Vanden-Eijnden, E.: Reconstruction of diffusions using spectral data from time-series. Commun. Math. Sci. 4, 651–668 (2006)
21. Crommelin, D., Vanden-Eijnden, E.: Diffusion estimation from multiscale data by operator eigenpairs. SIAM Multiscale Model. Simul. 9(4), 1588–1623 (2011)
22. Dacunha-Castelle, D., Dufo, M.: Probability and Statistics: Volume II. Springer, New York (1986)
23. Davies, E.B.: Heat Kernels and Spectral Theory. Cambridge University Press, Cambridge (1992)
24. Delattre, M., Genon-Catalot, V., Samson, A.: Estimation of population parameters in stochastic differential equations with random effects in the diffusion coefficient. MAP5 2014-07. HAL id: hal-01056917. to appear in ESAIM P&S. 2014
25. Ditljevson, S., Samson, A.: Estimation in the partially observed stochastic Morris-Lecar neuronal model with particle filter and stochastic approximation methods. Ann. Appl. Stat. 2, 674–702 (2014)
26. Dolapchis, S., Achatz, U., Timofeyev, I.: Stochastic closure for local averages in the finite-difference discretization of the forced Burgers equation. Theor. Comput. Fluid Dyn. 27(3–4), 297–317 (2013)
27. Eluerian, O., Chib, S., Shephard, N.: Likelihood inference for discretely observed nonlinear diffusions. Econometrica 69(4), 959–993 (2001)
28. Franzke, C., Majda, A.J.: Low-order stochastic mode reduction for a prototype atmospheric GCM. J. Atmos. Sci. 63, 457–479 (2006)
29. Franzke, C., Majda, A.J., Vanden-Eijnden, E.: Low-order stochastic mode reduction for a realistic barotropic model climate. J. Atmos. Sci. 62, 1722–1745 (2005)
30. Friedman, A.: Partial Differential Equations of Parabolic Type. Prentice-Hall, Englewood Cliffs (1964)
31. Friedrich, R., Siegert, S., Peinke, J., Lück, S., Siefert, M., Lindemann, M., Raethjen, G.D.J., Pfister, G.: Extracting model equations from experimental data. Phys. Lett. A 271, 217–222 (2000)
32. Gallant, R., Long, J.R.: Estimating stochastic differential equations efficiently by minimum chi-squared. Biometrika 84(1), 125–141 (1997)
33. Genon-Catalot, V.: Parameter estimation for stochastic differential equations from noisy observations. Maximum likelihood and filtering techniques. Lipari 2009 biomathematics summer school, MAP5 2010-03. HAL id: hal-00448996 (2010)
34. Genon-Catalot, V., Jeantheau, T., Laredo, C.: Limit theorems for discretely observed stochastic volatility models. Bernoulli 5(3), 283–303 (1998)
35. Genon-Catalot, V., Jeantheau, T., Laredo, C.: Parameter estimation for discretely observed stochastic volatility models. Bernoulli 5(5), 855–872 (1999)
36. Hairer, M.: Convergence of Markov Processes. Lecture Notes (2010)
37. Hansen, L., Scheinkman, J., Touzi, N.: Spectral methods for identifying scalar diffusions. J. Economet. 86, 1–32 (1998)
38. Heston, S.: A closed-form solution for options with stochastic volatility with applications to bond and currency options. Rev. Financ. Stud. 6(2), 327–343 (1993)
39. Heyde, C.C.: Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. Springer, New York (1997)
40. Il'in, A., Kalashnikov, A., Olejnik, O.: Linear equations of the second order of parabolic type. Russ. Math. Surv. 17(3), 1 (1962)
41. Gradisek, R.F.J., Siegert, S., Grabec, I.: Analysis of time series from stochastic processes. Phys. Rev. E 62, 3146–3155 (2000)
42. Jain, A., Timofeyev, I., Vanden-Eijnden, E.: Stochastic mode-reduction in models with conservative fast sub-systems. Comm. Math. Sci. 13(2), 297–314 (2015)
43. Karlin, S., Taylor, H.: A Second Course in Stochastic Processes. Academic Press, (1981)
44. Kelly, L., Platen, E., Sørensen, M.: Estimation for discretely observed diffusions using transform functions. J. Appl. Probab. 41A, 99–118 (2004)
45. Kessler, M., Sørensen, M.: Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli 5(2), 299–314 (1999)
46. Kutoyants, Y.A.: Statistical Inference for Ergodic Diffusion Processes. Springer, London (2004)
47. Ladishenskaya, O., Solonnikov, V., Uraltseva, N.: Linear and Quasi-linear Equations of Parabolic Type. Nauka, Moscow (1967)
48. Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: A mathematics framework for stochastic climate models. Commun. Pure Appl. Math. 54, 891–974 (2001)
49. Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: A priori tests of a stochastic mode reduction strategy. Phys. D 170, 206–252 (2002)
50. Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: Stochastic models for selected slow variables in large deterministic systems. Nonlinearity 19(4), 769–794 (2006)
51. Melbourne, I., Stuart, A.M.: A note on diffusion limits of chaotic skew-product flows. Nonlinearity 24, 1361 (2011)
52. Norris, J.R., Stroock, D.W.: Estimates on the fundamental solution to heat flows with uniformly elliptic coefficients. Proc. Lond. Math. Soc. 62, 373–402 (1991)
53. Panzar, L., van Zanten, H.: Nonparametric Bayesian inference for ergodic diffusions. J. Stat. Plan. Infer. 139(12), 4193–4199 (2009)
54. Papaspiliopoulos, O., Pokern, Y., Roberts, G.O., Stuart, A.M.: Nonparametric estimation of diffusions: a differential equations approach. Biometrika 99(3), 511–531 (2012)
55. Papavasiliou, A., Pavliotis, G.A., Stuart, A.: Maximum likelihood drift estimation for multiscale diffusions. Stoch. Proc. Appl. 119(10), 3173–3210 (2009)
56. Pavliotis, G.A., Stuart, A.: Parameter estimation for multiscale diffusions. J. Stat. Phys. 127, 741–781 (2007)
57. Pavliotis, G.A., Stuart, A.M.: Multiscale Methods: Averaging and Homogenization. Springer, New York (2008)
58. Pedersen, A.R.: A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Scand. J. Stat. 22(1), 55–71 (1995)
59. Pokern, Y., Stuart, A.M., van Zanten, J.H.: Posterior consistency via precision operators for Bayesian nonparametric drift estimation in SDEs. Stoch. Process. Appl. 123(2), 603–628 (2013)
60. Porper, F.O., Eidelman, S.D.: Two-sided estimates of fundamental solutions of second-order parabolic equations, and some applications. Russ. Math. Surv. 39(3), 119–178 (1984)
61. Prakasa Rao, B.L.S.: Statistical Inference for Diffusion Type Processes (Kendall’s Library Statist. 8). Arnold, London (1999)
62. Ren, P.: Parametric Estimation of the Heston Model Under the Indirect Observability Framework. PhD thesis, University of Houston, Dept. of Mathematics (2014)
63. Shepp, L.A., Marcus, M.B.: Sample behavior of Gaussian processes. In: Proc of 6th. Berkeley Symp., pp. 423–442 (1972)
64. Siegert, S., Friedrich, R., Peinke, J.: Analysis of data sets of stochastic systems. Phys. Lett. A 243, 275–280 (1998)
65. Sørensen, M.: Estimating Functions for Discretely Observed Diffusions: A Review. Lecture Notes–Monograph Series, vol. 32, pp. 305–326. Institute of Mathematical Statistics, Hayward (1997)
66. Stroock, D.W., Varadhan, S.: Multidimensional Diffusion Processes. Springer, New York (2005)
67. Sura, P., Barsugli, J.: A note on estimating drift and diffusion parameters from timeseries. Phys. Lett. A 305, 304–311 (2002)
68. Van Der Meulen, F.H., Van Der Vaart, A.W., Van Zanten, J.H.: Convergence rates of posterior distributions for Brownian semimartingale models. Bernoulli 12(5), 863–888 (2006)
69. Zanten, H.V.: Rates of convergence and asymptotic normality of kernel estimators for ergodic diffusion processes. J. Nonparametr. Stat. 13(6), 833–850 (2001)
70. Zhang, L., Mykland, P.A., Aït-Sahalia, Y.: A tale of two time scales: determining integrated volatility with noisy high-frequency data. J. Am. Stat. Assoc. 100(472), 1394–1411 (2005)