J. Stat. Phys. manuscript No.
(will be inserted by the editor)

R. Azencott · A. Beri · I. Timofeyev

Adaptive Sub-sampling for Parametric Estimation of Gaussian Diffusions

Received: date / Accepted: date

Abstract We consider a Gaussian diffusion X_t (Ornstein-Uhlenbeck process) with drift coefficient γ and diffusion coefficient σ^2, and an approximating process Y_t^ε converging to X_t in L_2 as $\varepsilon \to 0$. We study estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ which are asymptotically equivalent to the Maximum likelihood estimators of γ and σ^2, respectively. We assume that the estimators are based on the available $N = N(\varepsilon)$ observations extracted by sub-sampling only from the approximating process Y_t^ε with time step $\Delta = \Delta(\varepsilon)$. We characterize all such adaptive sub-sampling schemes for which $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ are consistent and asymptotically efficient estimators of γ and σ^2 as $\varepsilon \to 0$. The favorable adaptive sub-sampling schemes are identified by the conditions $\varepsilon \to 0$, $\Delta \to 0$, $(\Delta/\varepsilon) \to \infty$, and $N\Delta \to \infty$, which implies that we sample from the process Y_t^ε with a vanishing but coarse time step $\Delta(\varepsilon) >> \varepsilon$. This study highlights the necessity to sub-sample at adequate rates when the observations are not generated by the underlying stochastic model whose parameters are being estimated. The adequate sub-sampling rates we identify seem to retain their

R. Azencott
University of Houston
Department of Mathematics

Emeritus Professor
Ecole Normale Superieure, France
E-mail: razencot@math.uh.edu

A. Beri
University of Houston
Department of Mathematics
E-mail: arjunb@math.uh.edu

I. Timofeyev
University of Houston
Department of Mathematics
E-mail: ilya@math.uh.edu

validity in much wider contexts such as the additive triad application we briefly outline.

1 Introduction

Long-term evolution of high-dimensional deterministic systems governed by complex PDEs, have often been approximated by low-dimensional reduced stochastic models focused on larger time scales and with a good statistical fit to the observed dynamic data. For instance, stochastic mode-reduction technique [12–14] has successfully modeled the dynamics of large-scale structures in systems with time-scale separation, an optimal prediction setup has enabled coarse grain dynamic modeling of statistical descriptors [5,3,4], spin-flip processes have provided coarse-grained models of traffic flow [9–11,1], reduced Markov chain models have been applied to prototype atmosphere-ocean interactions [6], and a generic framework has been developed for dimension reduction in metastable systems [17,8]. In most practical contexts of this type, one seeks to approximate the dynamics of key statistical descriptors of a chaotic high-dimensional deterministic dynamical system by a closed form low-dimensional stochastic process, such as a (vector) stochastic differential equations (SDE). Then the available data $U_n = Y_{n\Delta}$, $n = 1, 2, ...N$ are not generated by the underlying SDE, but sampled from observations Y_t generated by some complex, not completely identifiable, deterministic dynamics.

On short time scales, the trajectories Y_t of the observable physical process are quite different from sample paths X_t of a vector SDE (see [7]), but on longer time-scales the behavior of Y_t is well emulated by X_t. This situation is typical for data generated by a numerical dynamic model, such as fluid dynamics PDEs. The Y_t trajectories then decorrelate slower than those of X_t, and good estimators $f(X_{t1}...X_{tN})$ of the underlying SDE parameters can lose their consistency if one simply substitutes X_t for Y_t in the function f and uses observations $(Y_{t1}...Y_{tN})$ which are too dense in time.

Sub-sampling strategies are then essential when the parameters of an SDE driven X_t must be estimated using discrete data extracted from a process Y_t^ε quite close to X_t for small ε, but having higher trajectory smoothness than X_t. Sub-sampling approaches have, for instance, been studied for the homogenization problem [16,15].

In this paper, for a class of Gaussian processes, we characterize efficient sub-sampling strategies with a complete determination of the optimal sub-sampling rates. For brevity, we focus on a prototypical case where X_t is driven by a one-dimensional Gaussian SDE and Y_t^ε is a Gaussian process with differentiable trajectories. Extension to higher-dimensional Gaussian diffusions looks quite feasible by similar techniques, and we expect to fully generalize to these cases our present characterization of good sub-sampling strategies enforcing consistent estimation of the underlying process parameters .

Our main results are presented via a prototype example in which X_t is a stationary Ornstein-Uhlenbeck (OU) process with unknown drift and diffusion coefficients γ and σ^2. We assume that the only available observations are generated by another stationary Gaussian process Y_t^ε, indexed by a small parameter $\varepsilon > 0$. We assume that as $\varepsilon \to 0$, the correlation function of Y_t^ε

converges to the correlation function of X_t. We analyze explicitly the case where Y_t^ε is generated by averaging X_t on a sliding time window $[t-\varepsilon, t]$. The Y_t^ε trajectories are then a.s. differentiable. The process X_t is not directly observable here, and the only available information is N observations extracted from Y_t^ε by *sub-sampling* with a time step Δ.

We consider estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ of γ and σ^2 based on the second-order sample moments of the underlying process. We show that these estimators are asymptotically equivalent to the maximum likelihood estimators (MLEs) $\hat{g}, \hat{s}$ of γ and σ^2. In these approximate MLEs $\hat{\gamma}$ and $\hat{\sigma}^2$, based on N unavailable observations of X_t, we replace the X_t by the available observables Y_t^ε, extracted by sub-sampling with a time step $\Delta > 0$. We study estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ for all *adaptive sub-sampling* schemes where ε determines the time step $\Delta = \Delta(\varepsilon)$ and the number of observations $N = N(\varepsilon)$. We show that if the *adaptive sub-sampling scheme* verifies,

$$\varepsilon \to 0, \quad \Delta \to 0, \quad (\Delta/\varepsilon) \to \infty, \quad N\Delta \to \infty,$$

then, the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ converge in L_2 to the underlying parameters γ, σ^2. Moreover, under the stronger conditions,

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad (N\varepsilon^2/\Delta) < cte,$$

where *cte* is an arbitrary positive constant, the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ converge with L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$.

The above conditions, as $\varepsilon \to 0$, provide an explicit recipe for the optimal choice of the time step $\Delta = \Delta(\varepsilon)$ and the number of observations $N = N(\varepsilon)$, given by,

$$\Delta \to 0, \quad (\Delta/\varepsilon) \to \infty, \quad (1/\Delta) << N < cte(\Delta/\varepsilon^2).$$

In the particular case where the number of observations N and the time step Δ are of the form $N = \varepsilon^{-\eta}$, $\Delta = \varepsilon^\alpha$, then, for α, η such that $\alpha \in (0,1)$, and $\alpha < \eta \leq (2-\alpha)$, the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ are asymptotically consistent estimators of γ, σ^2 with an L^2-speed of convergence proportional to $1/\sqrt{N\Delta}$. The best L_2-speeds of convergence are proportional to $\varepsilon^{1-\alpha}$, with α close to 0, and are reached for $\Delta = \varepsilon^\alpha$, $N = \varepsilon^{-2+\alpha}$.

Our key result is presented in Theorem (1). We have validated this result by numerical simulations presented in Section (9). A simple triad example is presented in Section (10) to illustrate the sub-sampling problem for a class of systems with the limiting behavior (as $\varepsilon \to 0$) given by the Ornstein-Uhlenbeck model. Our formalism is applicable in this case and can be used to analyze the consistency of the estimators. A simple intuitive explanation for the correct sub-sampling strategy is also provided. The triad example provided the initial motivation for investigating the sub-sampling problem and will be discussed in detail in a subsequent paper.

The outline of the paper is as follows. In sections (2) and (4), we present asymptotically efficient estimators, $\hat{\gamma}$ and $\hat{\sigma}^2$, for the OU parameters of the stationary OU process and prove their asymptotic equivalence to the MLEs. We apply known asymptotic results to *fixed rate sub-sampling* where observations are extracted from the OU trajectories themselves. We extend the consistency results to *adaptive sub-sampling* schemes in section (5).

In section (6) we present the Smoothed OU process Y_t^ε and *indirect estimators* of OU parameters based on observations extracted from Y_t^ε. In sections (7) and (8), we characterize the optimal *adaptive sub-sampling* for such *indirect estimators* and discuss the properties of the estimators. In section (9) we illustrate numerically various sub-sampling strategies on simulated diffusions. In Section (10) we briefly outline the additive triad application and discuss numerical results obtained by using various adaptive sub-sampling strategies for observations from the triad model.

2 Ornstein-Uhlenbeck Process

Consider a filtered probability space $(\Omega, F, P; \mathbb{F})$ on which we define the stochastic processes of interest. In dimension 1, we consider a basic example of Gaussian diffusion, namely the Ornstein-Uhlenbeck process (denoted here as *OU-process*), defined as the solution for $t \geq 0$ of the linear SDE

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{1}$$

where W_t is the standard Brownian motion and the unknown parameters $\gamma; \sigma$ are strictly positive. The solution X_t of SDE (1) is an asymptotically stationary Gaussian process given by

$$X_t = X_0 e^{-\gamma t} + \sigma e^{-\gamma t} \int_0^t e^{\gamma s} dW_s. \tag{2}$$

When $X_0 = x_0 \in \mathbb{R}$, the distribution of X_t is given by

$$X_t \sim \mathbf{N}\left(X_0 e^{-\gamma t}, \frac{\sigma^2}{2\gamma}(1 - e^{-2\gamma t})\right), \tag{3}$$

where $\mathbf{N}(a, b)$ is the Gaussian distribution with mean a and variance b. The covariance function of $(X_t)_{t \geq 0}$ is given by

$$E[X_t X_s] = \frac{\sigma^2}{2\gamma} e^{-\gamma|t-s|} \left(1 - e^{-2\gamma(s \wedge t)}\right) + X_0^2 e^{-\gamma(t+s)},$$

where $(s \wedge t) = min\{s, t\}$. Since $\gamma > 0$, the process $(X_t)_{t \geq 0}$ is asymptotically stationary as $t \to \infty$, and converges in distribution to the Gaussian $\mathbf{N}\left(0, \left(\sigma^2/2\gamma\right)\right)$. The asymptotic stationary covariance is given by

$$\lim_{t \to \infty} E[X_t X_{t+h}] = \frac{\sigma^2}{2\gamma} e^{-\gamma|h|}. \tag{4}$$

When $X_0 \sim \mathbf{N}\left(0, \sigma^2/2\gamma\right)$, then X_t is a stationary process. In this case, the asymptotic results for parametric estimation of stationary Gaussian processes from [2] are directly applicable to the discretely sub-sampled OU process. These results, which we recall below, are easily extended to the situation where the OU process is only asymptotically stationary because convergence to its stationary distribution is exponentially fast. Hence, the OU process observed for $t \geq t_0$ such that $t_0 >> (\gamma^{-1})$, may essentially be considered as stationary.

We now analyze the asymptotic properties of maximum likelihood estimators (denoted here MLEs) of the unknown parameters γ and σ based on large sets of sub-sampled but direct observations of the process X_t. These MLEs are differentiable functions of the covariance estimators of the OU process, a representation which is crucial here.

3 Fixed Rate sub-sampling and Adaptive sub-sampling

Assume $N \to \infty$, and that we have $(N+1)$ direct observations $U_n = X_{n\Delta}$ with $n = 0, ..., N$, extracted from the OU-trajectory X_t by *sub-sampling at discrete time steps* $t = n\Delta$.

Definition 1 We say that we have a *Fixed rate Sub-sampling* scheme when the time-step $\Delta > 0$ between observations remains fixed as $N \to \infty$.

□

Definition 2 We say that we have an *Adaptive Sub-sampling* scheme when the time-step between observations depends on N, i.e. $\Delta = \Delta(N) \to 0$ as $N \to \infty$, and we then *always impose the condition* $N\Delta \to \infty$.

□

As shown below, when the global time interval $N\Delta$ spanned by the N available observations remains bounded, the maximum likelihood estimators of γ, σ based on these N observations are not be asymptotically consistent. This is due to the $O(1/\sqrt{N\Delta})^m$, $m \geq 1$, bias terms in the asymptotic expansions of the estimators about the true values.

From (2) and (3), we infer that the U_n satisfy the difference equation,

$$U_{n+1} = U_n e^{-\gamma\Delta} + \sqrt{\frac{\sigma^2(1 - e^{-2\gamma\Delta})}{2\gamma}} Z_n, \tag{5}$$

where the Z_n are i.i.d. standard Gaussian variables, and are independent of $U_0, ..., U_n$, for each $n = 0, \ldots, N-1$. When $U_0 = X_0 \sim \mathbb{N}\left(0, \frac{\sigma^2}{2\gamma}\right)$, then (U_n) is a Gaussian stationary Markov process.

Proposition 1 *Let $(U_n)_{n\in\mathbb{Z}}$ be a centered stationary Gaussian process. Define the covariances r_k for each $k \in \mathbb{Z}$ by*

$$r_k = E\left[U_n U_{n+k}\right].$$

Define the empirical covariance estimators $\hat{r}_k(N)$ by

$$\hat{r}_k(N) = (1/N) \sum_{n=0}^{N-1} U_n U_{n+k}.$$

Then for each pair of non-negative integers k, q, the covariance of the estimators $C_{kq} = Cov(\hat{r}_k(N), \hat{r}_q(N))$ is given by

$$C_{kq} = (1/N) \sum_{j=-(N-1)}^{N-1} f(j) - (1/N^2) \sum_{j=1}^{N-1} j\,(f(j) + f(-j))\,, \tag{6}$$

where $f(j) = r_j r_{j+k-q} + r_{j+k} r_{j-q}$.

Proof . The covariance $Cov(\hat{r}_k(N), \hat{r}_q(N)) = E\left[\hat{r}_k(N)\hat{r}_q(N)\right] - r_k r_q$, where $E\left[\hat{r}_k(N)\hat{r}_q(N)\right]$ can be explicitly computed from the 4th order moments of a Gaussian random vector, and is given by

$$N^2 E\left[\hat{r}_k(N)\hat{r}_q(N)\right] = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E\left[U_m U_{m+k} U_n U_{n+q}\right]. \tag{7}$$

A well known result for the Gaussian random variables gives us the 4th order moments in terms of the 2nd order moments, namely

$$E\left[U_m U_{m+k} U_n U_{n+q}\right] = r_k r_q + r_{m-n} r_{m-n+k-q} + r_{m-n-q} r_{m-n+k}. \tag{8}$$

Substituting (8) in (7) gives the required result. □

4 Fixed Rate Sub-sampling

Assume *fixed-rate sub-sampling*, so that $\Delta > 0$ is fixed. The stationary covariances r_k are given by

$$r_k = r_k(\Delta) = E[U_{n+k} U_n] = \frac{\sigma^2}{2\gamma} e^{-\gamma k \Delta}, \tag{9}$$

which implies the relation

$$\sum_{k \in \mathbb{Z}} |k| |r_k(\Delta)| < \infty. \tag{10}$$

Given the discrete observations U_n, define the standard *empirical covariance estimators* $(\hat{r}_k)_{k=0,1}$ by

$$\hat{r}_0 = \hat{r}_0(N, \Delta) = \frac{1}{N} \sum_{n=0}^{N-1} U_n^2, \qquad \hat{r}_1 = \hat{r}_1(N, \Delta) = \frac{1}{N} \sum_{n=0}^{N-1} U_{n+1} U_n. \tag{11}$$

Since the covariances r_k verify (10), known results [2] on stationary Gaussian processes show that for each fixed $\Delta > 0$ as $N \to \infty$, the covariance estimators $\hat{r}_k$ are the best estimators of the r_k, they are *consistent* (i.e. converges almost surely to the true r_k), and *asymptotically efficient* (i.e. the asymptotic variances of $\hat{r}_k$ attain the Cramér-Rao bound). We also know (see [2, Chapter X]) that as $N \to \infty$, the random vectors

$$\sqrt{N}\left[\hat{r}_0 - r_0, \hat{r}_1 - r_1\right]$$

are asymptotically centered and Gaussian, with limit covariance matrix $\Gamma = (\Gamma_{st})$, $s, t \in \{0;1\}$ given by

$$\Gamma_{st} = \sum_{m\in\mathbb{Z}} \left(r_m r_{m-s+t} + r_{m-s} r_{m+t} \right),$$

with r_m given by (9), and hence, the covariance matrix Γ is given by

$$\begin{aligned} \Gamma_{00} &= 2r_0^2(1+e^{-2\gamma\Delta}) / \left(1-e^{-2\gamma\Delta}\right), \\ \Gamma_{11} &= r_0^2\left(1+4e^{-2\gamma\Delta}-e^{-4\gamma\Delta}\right) / \left(1-e^{-2\gamma\Delta}\right), \\ \Gamma_{01} &= \Gamma_{10} = 4r_0^2 e^{-\gamma\Delta} / \left(1-e^{-2\gamma\Delta}\right), \end{aligned} \tag{12}$$

where $r_0 = \dfrac{\sigma^2}{2\gamma}$. The relation (9) between OU parameters γ, σ^2 and the covariances r_0 and r_1 imply that

$$\gamma = g(r_0, r_1), \qquad \sigma^2 = s(r_0, r_1),$$

where the smooth functions g, s are given by

$$g(r_0, r_1) = (-1/\Delta) \log\left(\frac{r_1}{r_0}\right), \qquad s(r_0, r_1) = (-2r_0/\Delta) \log\left(\frac{r_1}{r_0}\right).$$

We now study the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ for γ and σ^2 given by

$$\hat{\gamma} = g(\hat{r}_0, \hat{r}_1), \qquad \hat{\sigma}^2 = s(\hat{r}_0, \hat{r}_1),$$

which have the explicit expressions

$$\hat{\gamma} = -\frac{1}{\Delta} \log\left(\frac{\hat{r}_1}{\hat{r}_0}\right), \qquad \hat{\sigma}^2 = 2\hat{\gamma}\hat{r}_0. \tag{13}$$

Proposition 2 (Asymptotics for $\hat{\gamma}$ and $\hat{\sigma}^2$) *Consider an OU-process X_t directly observed at times $t = n\Delta$, $n = 0, \dots, N$, sub-sampling at fixed rate $\Delta > 0$. Then as $N \to \infty$, the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ of γ and σ^2 are consistent (almost surely). Moreover $\sqrt{N}(\hat{\gamma} - \gamma)$ and $\sqrt{N}(\hat{\sigma}^2 - \sigma^2)$ are asymptotically Gaussian with limit variances v_γ and v_{σ^2} given by*

$$\begin{aligned} v_\gamma &= \left(\frac{e^{-2\gamma\Delta}+e^{2\gamma\Delta}-2}{\Delta^2(1-e^{-2\gamma\Delta})}\right), \\ v_{\sigma^2} &= 4r_0^2\left(\frac{2(1+\gamma\Delta)^2(1+e^{-2\gamma\Delta})-8\gamma\Delta+e^{2\gamma\Delta}-e^{-2\gamma\Delta}-4}{\Delta^2(1-e^{-2\gamma\Delta})}\right). \end{aligned}$$

Proof . Define the function $F: \mathbb{R}^2 \to \mathbb{R}^2$ by

$$F(r_0, r_1) = [g(r_0, r_1), s(r_0, r_1)].$$

Since F is twice continuously differentiable in the neighborhood of (r_0, r_1) for each $\Delta > 0$, it follows from [2, Chapter X] that the estimator $\hat{\nu} = F(\hat{r}_0, \hat{r}_1)$ is a consistent estimator of $\nu = F(r_0, r_1)$. Also, the distribution of $\sqrt{N}\,(\hat{\nu} - \nu)$

converges, as $N \to \infty$, toward the 2-dimensional centered Gaussian distribution with covariance matrix

$$\Sigma_F = A \Gamma A^T$$

where, for each fixed $\Delta > 0$, the (2×2) matrix $A = \nabla F(r_0, r_1)$ is the differential of F at true covariances (r_0, r_1), A^T denotes the transpose of A, and the covariance matrix Γ is given by (12). This says exactly that $\hat{\gamma}$ and $\hat{\sigma}^2$ are *consistent and asymptotically Gaussian estimators* of γ and σ^2, and that Σ_F is the limit covariance matrix of the random vector

$$\sqrt{N}\left[(\hat{\gamma} - \gamma), \left(\hat{\sigma}^2 - \sigma^2\right)\right].$$

□

Recall that 2 asymptotically Gaussian estimators $\tau_{1,N}$ and $\tau_{2,N}$ of a parameter τ are said to be *asymptotically equivalent* if $\sqrt{N}(\tau_{1,N} - \tau)$ and $\sqrt{N}(\tau_{2,N} - \tau)$ have the same limit variance as $N \to \infty$.

Proposition 3 (MLEs) *The log likelihood $L_{N,\Delta}$ of $U = \{U_n\}$ is given by*

$$L_{N,\Delta}(U; \gamma, \sigma^2) = -\frac{1}{2} \log\left(\frac{\pi\sigma^2}{\gamma}\right) - \frac{\gamma U_0^2}{\sigma^2} - \frac{N}{2} \log\left(\frac{\pi\sigma^2}{\gamma}(1 - e^{-2\gamma\Delta})\right)$$

$$-\gamma\left(\sigma^2\left(1 - e^{-2\gamma\Delta}\right)\right)^{-1} \sum_{n=0}^{N-1} (U_{n+1} - e^{-\gamma\Delta} U_n)^2.$$

Call $\hat{g} = \hat{g}(N, \Delta)$ and $\hat{s} = \hat{s}(N, \Delta)$ the maximum likelihood estimators of γ and σ^2, obtained by maximizing $L_{N,\Delta}$ in γ, σ^2. Then, the MLE estimators $\hat{g}$ and $\hat{s}$ are resp. asymptotically equivalent to the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ defined above by (13).

Proof . From (5), for a fixed $\Delta > 0$, we derive the Gaussian conditional density of U_{n+1} given U_n, and then the likelihood $exp(-L_{N,\Delta})$ of $\{U_0, \ldots, U_N\}$ using the Markov property of the U_n. This yields the explicit expression of $L_{N,\Delta}$ stated above.

The values $\left(\gamma, \sigma^2\right)$ maximizing $L_{N,\Delta}$ must verify the necessary conditions $\nabla L_{N,\Delta}(\gamma, \sigma^2) = 0$, namely the two equations

$$\sigma^2 = \left(\frac{2\gamma}{N+1}\right)\left(N\hat{r}_0 + \frac{2\gamma\left(U_N^2 - U_0^2 e^{-2\gamma\Delta}\right)}{(1 - e^{-2\gamma\Delta})}\right), \tag{14}$$

$$\gamma = \frac{-1}{\Delta} \log\left(\frac{\hat{r}_1}{\hat{r}_0 + N^{-1}\left((\sigma^2/2\gamma) - U_0^2\right)}\right).$$

The Hessian of $L_{N,\Delta}$ is negative definite for large N, hence for large N, $L_{N,\Delta}$ has a unique supremum reached at the point $(\hat{g}, \hat{s})$, which solves equations (14). Note that these equations do not lead to an explicit expression for $(\hat{g}, \hat{s})$. Using large deviation bounds for Gaussian processes, one proves that with very large probability, $\hat{\gamma}$ and $\hat{\sigma}^2$ is a good approximation to the solution of (14), with an accuracy of the order of $\dfrac{1}{\sqrt{N}}$ as $N \to \infty$. □

We have seen that under *fixed-rate sub-sampling scheme* the covariance estimators $\hat{r}_k$ and the OU estimators $\hat{\gamma}$, $\hat{\sigma}^2$ are *consistent* (in L_2) and *asymptotically Gaussian*. The L_2-speed of convergence for the OU estimators $\hat{\gamma}$, $\hat{\sigma}^2$ are proportional to $1/\sqrt{N}$, for each fixed value of $\Delta > 0$.

5 Adaptive Sub-sampling

We now study the consistency of the estimators $\hat{\gamma} = \hat{\gamma}(N, \Delta)$ and $\hat{\sigma}^2 = \hat{\sigma}^2(N, \Delta)$ under *adaptive sub-sampling scheme* (see definition (2)).

Proposition 4 (Asymptotics of the Covariances) *Consider an adaptive sub-sampling scheme where we have N observations $U_n = X_{n\Delta}$ of the stationary OU process X_t at time intervals of length $\Delta = \Delta(N)$ depending on N. We assume (see definition (2))*

$$\Delta \to 0, \quad N\Delta \to \infty. \tag{15}$$

The true covariances $r_k = r_k(\Delta)$ of the process U_n are now functions of N still given by (9)*. Hence as $N \to \infty$, and for each $k \geq 0$, $r_k(\Delta(N)) \to (\sigma^2/2\gamma)$.*

Then, under condition (15)*, and for each $k \geq 0$, the empirical covariances $\hat{r}_k$ converge in L_2 to $(\sigma^2/2\gamma)$. Moreover, for each $k \geq 0$ the L^2-norms of the variables $\sqrt{N\Delta}\,(\hat{r}_k - r_k)$ converge to $\left(\sigma^2/\gamma\sqrt{2\gamma}\right)$.*

Proof . The associated speed of convergence to zero in L_2 for the difference $(\hat{r}_k - r_k)$ can be computed directly, as outlined here for $k = 0$ and $k = 1$. Let $J_k = E\left[(\hat{r}_k - r_k)^2\right]$, then, using the expression (6) in proposition (1), such that $J_k = C_{kk}$, we obtain for $k = 0, 1$,

$$\begin{aligned} J_0 &= \frac{2r_0^2(1 + e^{2\gamma\Delta})}{N(e^{2\gamma\Delta} - 1)} + \frac{4r_0^2 e^{2\gamma\Delta}\left(e^{-2\gamma N\Delta} - 1\right)}{N^2(e^{2\gamma\Delta} - 1)^2}, \\ J_1 &= \frac{r_0^2(e^{4\gamma\Delta} + 4e^{2\gamma\Delta} - 1)}{Ne^{2\gamma\Delta}(e^{2\gamma\Delta} - 1)} + \frac{4r_0^2 e^{2\gamma\Delta}\left(e^{-2\gamma N\Delta} - 1\right)}{N^2(e^{2\gamma\Delta} - 1)^2}. \end{aligned} \tag{16}$$

Under the conditions (15), and as $\Delta \to 0, N\Delta \to \infty$ using the convergence of $\left(\Delta/(e^{2\gamma\Delta} - 1)\right) \to (2\gamma)^{-1}$, and $N(e^{2\gamma\Delta} - 1) \to \infty$ (for $\gamma > 0$), we have,

$$(N\Delta)J_0 \to \frac{\sigma^4}{2\gamma^3}, \qquad (N\Delta)J_1 \to \frac{\sigma^4}{2\gamma^3}.$$

This concludes the proof. □

Proposition 5 *For each N, Δ, the random variables $Z_0 = Z_0(N, \Delta)$ and $Z_1 = Z_1(N, \Delta)$ defined by,*

$$Z_0 = \frac{(\hat{r}_0 - r_0)}{\sqrt{J_0}}, \qquad Z_1 = \frac{(\hat{r}_1 - r_1)}{\sqrt{J_1}}, \tag{17}$$

have mean 0, variance 1, and covariance $E[Z_0 Z_1] = J_{01}/(\sqrt{J_0 J_1})$, *where* J_0, J_1, J_{01} *are given by* (16), (19). *Then, under conditions* (15) *we have the following first-order* L_2 *approximations for the empirical covariances* $\hat{r}_k$,

$$\begin{aligned} \hat{r}_0 &= r_0 + \frac{r_0}{\sqrt{N\Delta}}\sqrt{\frac{2}{\gamma}} Z_0 + \frac{Z_0}{\sqrt{N\Delta}} \times O\left(\Delta^2 + \frac{1}{N\Delta}\right), \\ \hat{r}_1 &= r_1 + \frac{r_0}{\sqrt{N\Delta}}\sqrt{\frac{2}{\gamma}} Z_1 + \frac{Z_1}{\sqrt{N\Delta}} \times O\left(\Delta^2 + \frac{1}{N\Delta}\right), \end{aligned} \tag{18}$$

where the notation $O(h)$ *represents deterministic functions of* h *bounded by a constant multiple of* h.

Proof . The exact expression for J_{01} derived using proposition (1) is given by,

$$J_{01} = \frac{4r_0^2 e^{\gamma\Delta}}{N(e^{2\gamma\Delta}-1)} - \frac{2r_0^2 e^{\gamma\Delta}(e^{2\gamma\Delta}+1)(1-e^{-2\gamma N\Delta})}{N^2(e^{2\gamma\Delta}-1)^2}. \tag{19}$$

Applying Taylor's expansions to J_k as given by (16) and J_{01} given by (19), we obtain

$$\begin{aligned} J_0 &= \frac{2r_0^2}{\gamma N\Delta}\left(1 + \frac{\gamma^2\Delta^2}{3} - \frac{1+O(\Delta^2)}{2\gamma N\Delta} + O(\Delta^4)\right), \\ J_1 &= \frac{2r_0^2}{\gamma N\Delta}\left(1 - \frac{2\gamma^2\Delta^2}{3} + \gamma^3\Delta^3 - \frac{1+O(\Delta^2)}{2\gamma N\Delta} + O(\Delta^4)\right), \\ J_{01} &= \frac{2r_0^2}{\gamma N\Delta}\left(1 - \frac{\gamma^2\Delta^2}{6} - \frac{1+O(\Delta^2)}{2\gamma N\Delta} + O(\Delta^4)\right). \end{aligned} \tag{20}$$

Substituting in (17) the above expressions for J_0, J_1 gives the required L_2-approximations as expressed in (18). □

Define the random variable Z_k, for any integer $k \geq 0$, as $Z_k = (\hat{r}_k - r_k)/\sqrt{J_k}$, where $J_k = C_{kk}$ is given by (6), in particular for the OU process. The next lemma will be needed to prove the consistency of $\hat{\gamma}$ and $\hat{\sigma}^2$.

Lemma 1 *For each integer* $k \geq 0$, *consider a random variable* $V_k = V_k(\theta)$ *given by,*

$$V_k = \left(\frac{a_k Z_k}{1 + a_k \theta Z_k}\right)^2, \tag{21}$$

where $Z_k = (\hat{r}_k - r_k)/\sqrt{J_k}$, $\theta \in (0,1)$ *and* $a_k = e^{\gamma k\Delta}\sqrt{J_k}/r_0$, *such that* $J_k \sim O(1/N\Delta)$. *Then, under the condition* (15), $\|V_k\|_{L_2} \to 0$, *at a speed proportional to* $1/N\Delta$.

Proof . The L^2-norm is given by $\|V_k\|_{L_2}^2 = E\left[(a_k Z_k/(1 + a_k\theta Z_k))^4\right]$. Since, the tails of the density for Z_k decay exponentially fast, we have for any

$M \gg 1$, $P\left\{(1+a_k\theta Z_k)^{-1} > M\right\} < e^{-(C_1\sqrt{N\Delta}/\theta)}$, where C_1 is a positive constant. Also,

$$P\left\{(1+a_k\theta Z_k)^{-1} < 0\right\} = P\left\{Z_k < -(C_2\sqrt{N\Delta}/\theta)\right\} < e^{-(C_2/\theta)\sqrt{N\Delta}},$$

where C_2 is a positive constant. Therefore, using Cauchy-Schwarz inequality we obtain,

$$\|V_k\|_{L^2}^2 \leq a_k^4 \|Z_k^4\|_{L_2} \|(1+a_k\theta Z_k)^{-4}\|_{L_2} \leq \frac{C_3}{N^2\Delta^2},$$

where $\|Z_k^4\|_{L_2}$ is uniformly bounded in N, Δ for each k and C_3 is some positive constant. This proves the required result. □

Proposition 6 (Consistency of $\hat{\gamma}$ and $\hat{\sigma}^2$) *Consider the adaptive subsampling scheme providing N observations $U_n = X_{n\Delta}$ of the stationary OU process X_t at time intervals of length $\Delta = \Delta(N)$. Define the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ by formula* (13).

Then, under the conditions

$$\Delta \to 0, \quad N\Delta \to \infty, \tag{22}$$

the estimators $\hat{\gamma}$, $\hat{\sigma}^2$ are asymptotically consistent estimators of γ, σ^2, i.e., $\hat{\gamma} \to \gamma$, and $\hat{\sigma}^2 \to \sigma^2$ in L_2.

Moreover, given (22), *the L_2 norms of the variables $\sqrt{N\Delta}\,(\hat{\gamma}-\gamma)$, and $\sqrt{N\Delta}\left(\hat{\sigma}^2-\sigma^2\right)$ converge, respectively, to $\sqrt{2\gamma}$ and 0. Therefore, the estimators converge to the true values with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$. In particular, under stronger conditions,*

$$\Delta \to 0, \quad N\Delta^2 \to \infty, \tag{23}$$

the L_2-speed of convergence of the estimator $\hat{\sigma}^2$ to σ^2 is proportional to $1/\sqrt{N}$, such that $\|\sqrt{N}\left(\hat{\sigma}^2-\sigma^2\right)\|_{L_2} \to \sigma^2\sqrt{2}$.

Proof . From (17) we obtain, $\hat{r}_0 = r_0 + \sqrt{J_0}Z_0$, and $\hat{r}_1 = r_1(\Delta) + \sqrt{J_1}Z_1$, which we substitute in

$$\hat{\gamma} = -\frac{1}{\Delta}\log\left(\frac{\hat{r}_1}{\hat{r}_0}\right).$$

First, we rewrite the ratio $\hat{R} = (\hat{r}_1/\hat{r}_0)$ as follows,

$$\hat{R} = e^{-\gamma\Delta}\left(1+\frac{e^{\gamma\Delta}\sqrt{J_1}}{r_0}Z_1\right)\left(1+\frac{\sqrt{J_0}}{r_0}Z_0\right)^{-1}.$$

Then taking logarithms of the ratio $\hat{R}$, we obtain,

$$\hat{\gamma} = \gamma - \frac{1}{\Delta}\log\left(1+\frac{e^{\gamma\Delta}\sqrt{J_1}}{r_0}Z_1\right) + \frac{1}{\Delta}\log\left(1+\frac{\sqrt{J_0}}{r_0}Z_0\right).$$

Using lemma(1), under the conditions (22), and using Taylor's expansion we obtain that the following holds in L_2,

$$\begin{aligned}\log\left(1+\frac{\sqrt{J_0}}{r_0}Z_0\right) &= \frac{\sqrt{J_0}}{r_0}Z_0-(V_0/2),\\ \log\left(1+\frac{e^{\gamma\Delta}\sqrt{J_1}}{r_0}Z_1\right) &= \frac{e^{\gamma\Delta}\sqrt{J_1}}{r_0}Z_1-(V_1/2),\end{aligned}$$

where the random remainder terms V_0, V_1 are given by (21) such that the L_2 norms $\|V_1-V_0\|_{L_2}\sim O(1/N)$, $\|V_0\|_{L_2}\sim O(1/N\Delta)$, and $\|V_1\|_{L_2}\sim O(1/N\Delta)$.

Let the random variable $Z_\gamma = \left(e^{\gamma\Delta}\sqrt{J_1}Z_1-\sqrt{J_0}Z_0\right)/(\Delta r_0)$. The L_2 norm of Z_γ is given by,

$$\|Z_\gamma\|^2_{L^2} = \frac{\left(e^{2\gamma\Delta}J_1+J_0-2e^{\gamma\Delta}J_{01}\right)}{(\Delta r_0)^2} = \frac{2\gamma}{N\Delta}\left(1+O(\Delta)\right).$$

Then, the first-order L_2 approximation for $\hat{\gamma}$ is given by,

$$\hat{\gamma} = \gamma - Z_\gamma + R_\gamma \times O\left(\frac{1}{N\Delta}\right), \tag{24}$$

where the random remainder term $R_\gamma = R_\gamma(N,\Delta)$ is uniformly bounded in L_2 norm. Therefore, under the conditions (22), the estimator $\hat{\gamma}\to\gamma$ in L_2 with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$ such that,

$$\|\sqrt{N\Delta}\,(\hat{\gamma}-\gamma)\,\|_{L^2}\to\sqrt{2\gamma}.$$

The diffusion estimator $\hat{\sigma}^2 = 2\hat{\gamma}\hat{r}_0$ by (13). Let the random variable $Z_{\sigma^2} = (2/\Delta)\left(e^{\gamma\Delta}\sqrt{J_1}Z_1-(1+\gamma\Delta)\sqrt{J_0}Z_0\right)$, then its L_2 norm is given by,

$$\|Z_{\sigma^2}\|^2_{L_2} = \frac{2\sigma^4}{N}\left(1+O(\Delta)\right).$$

Hence, using (17) and (24) we obtain,

$$\hat{\sigma}^2 = \sigma^2 - Z_{\sigma^2} + R_{\sigma^2}\times O\left(\frac{1}{N\Delta}\right), \tag{25}$$

where the random remainder term $R_{\sigma^2} = R_{\sigma^2}(N,\Delta)$ is uniformly bounded in L_2 norm. Therefore, under the conditions (22), $\hat{\sigma}^2\to\sigma^2$ in L_2. Moreover, under the conditions (23), the following convergence holds,

$$\|\sqrt{N}\left(\hat{\sigma}^2-\sigma^2\right)\|_{L^2}\to\sigma^2\sqrt{2}.$$

□

To summarize, when the observations are directly extracted from a stationary OU process then, under the *fixed rate sub-sampling scheme* the MLEs for the parameters of the OU-process are consistent and asymptotically Gaussian. The L_2-speed of convergence for the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ as $N \to \infty$ is proportional to $1/\sqrt{N}$ for each fixed $\Delta > 0$.

Under the *adaptive sub-sampling scheme* (22), the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ are asymptotically consistent estimators of γ, σ^2. The usual L_2-speed of convergence to true values proportional to $1/\sqrt{N\Delta}$ is achievable for the estimators $\hat{\gamma}$, $\hat{\sigma}^2$. In fact for the diffusion estimator $\hat{\sigma}^2$, under stronger conditions on N, Δ, one can achieve a faster L_2-speed of convergence proportional to $1/\sqrt{N}$.

The asymptotic distribution of the empirical covariance estimators and the OU estimators, under the *adaptive sub-sampling scheme*, will be studied elsewhere.

We now study a more common and more complex scenario in which only indirect observations of the underlying OU-process X_t are available, and are generated by another process Y_t which is not identical to X_t, but is simply close to X_t in L_2. In this case *sub-sampling will become an essential tool to generate consistent estimators* of the underlying parameters.

6 Indirect Estimation of OU-Parameters

Assume now that the stationary OU process X_t is *not directly observable*, and that all available observations are extracted from a centered stationary process Y_t^ε, which tends to the process X_t in L_2 as $\varepsilon \to 0$. More precisely, defining the covariance functions

$$K^\varepsilon(h) = E[Y_t^\varepsilon Y_{t+h}^\varepsilon], \quad \text{and} \quad r_h = E[X_t X_{t+h}],$$

we assume that

$$K^\varepsilon(h) \to r_h \quad \text{as} \quad \varepsilon \to 0.$$

We focus here on one precise example of this situation, namely, the specific case where the process Y_t^ε is the *Smoothed Ornstein-Uhlenbeck* process, also denoted SOU-process, obtained by averaging the OU process over a sliding window of fixed length $\varepsilon > 0$, so that

$$Y_t^\varepsilon = \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} X_s ds. \tag{26}$$

Note that Y_t^ε is a centered stationary Gaussian process with a.s. differentiable trajectories. The covariance function of Y_t^ε at time lag h is given by,

$$K^\varepsilon(h) = E[Y_t^\varepsilon Y_{t+h}^\varepsilon] = \frac{1}{\varepsilon}\left(\int_{t+h-\varepsilon}^{t+h} E[X_s Y_t^\varepsilon] ds\right).$$

As is well known, we may in this Gaussian context freely commute expectation signs and integral signs, so that the computation of $K^\varepsilon(h)$ boils down to computing simple deterministic integrals of the explicit stationary covariance function of X_t. We thus obtain the following expressions for $K^\varepsilon(h)$, for $h \geq 0$:

$$K^\varepsilon(h) = \begin{cases} \dfrac{\sigma^2}{2\gamma^3\varepsilon^2} e^{-\gamma h} \left(e^{-\gamma\varepsilon} + e^{\gamma\varepsilon} - 2\right). & h \geq \varepsilon, \\ \\ \dfrac{\sigma^2}{2\gamma^3\varepsilon^2} e^{-\gamma h} \big(2\gamma(\varepsilon - h)e^{\gamma h} + e^{-\gamma\varepsilon}(e^{2\gamma h} + 1) - 2\big). & h < \varepsilon. \end{cases} \tag{27}$$

In particular, we have

$$K^\varepsilon(0) = \frac{\sigma^2}{\gamma^3\varepsilon^2}(e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon). \tag{28}$$

Therefore, the correlation function of Y^ε is given by

$$\frac{K^\varepsilon(h)}{K^\varepsilon(0)} = \frac{1}{2} e^{-\gamma h} \left(\frac{e^{-\gamma\varepsilon} + e^{\gamma\varepsilon} - 2}{e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon}\right), \quad h \geq \varepsilon, \tag{29}$$

$$\frac{K^\varepsilon(h)}{K^\varepsilon(0)} = \frac{1}{2} e^{-\gamma h} \left(\frac{2\gamma(\varepsilon - h)e^{\gamma h} + e^{-\gamma\varepsilon}(e^{2\gamma h} + 1) - 2}{e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon}\right), \quad 0 \leq h < \varepsilon. \tag{30}$$

Next we present the study of the asymptotic properties of the estimators, $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$, based on the observations sub-sampled from the process Y^ε_t.

7 Fixed Rate Sub-sampling for Indirect Estimation of parameters

Recall that now the *only available information* are (N+1) *indirect* observations $U^\varepsilon_n = Y^\varepsilon_{n\Delta}$ extracted from the SOU-process Y^ε_t by sub-sampling with a fixed time-step $\Delta > 0$. The goal is still to estimate the parameters γ and σ^2 of the underlying OU process. We will study the estimators given by formulas (13) where we replace U_n by U^ε_n. These approximate MLEs of γ and σ^2, are given by

$$\hat{\gamma}_\varepsilon = -\frac{1}{\Delta} \log\left(\frac{\hat{r}^\epsilon_1}{\hat{r}^\epsilon_0}\right), \quad \hat{\sigma}^2_\varepsilon = 2\hat{\gamma}_\varepsilon \hat{r}^\epsilon_0, \tag{31}$$

where $\hat{r}^\epsilon_k = (1/N)\sum_{n=0}^{N-1} U^\varepsilon_n U^\varepsilon_{n+k}$ is the standard empirical estimator of the covariance $r^\varepsilon_k = K^\varepsilon(k\Delta)$ given by (27), for $k = 0, 1$.

Proposition 7 (Asymptotic Bias of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$) *For fixed ε and Δ the following convergence holds in L_2 as $N \to \infty$, namely,*

$$\hat{\gamma}_\varepsilon \to G = G(\varepsilon, \Delta), \qquad \hat{\sigma}^2_\varepsilon \to S = S(\varepsilon, \Delta),$$

where

$$G = -(1/\Delta)\log\left(K^\varepsilon(\Delta)/K^\varepsilon(0)\right) \quad \text{and} \quad S = 2GK^\varepsilon(0), \tag{32}$$

and where the covariances $K^\varepsilon(0)$ and $K^\varepsilon(\Delta)$ are given by (27).

Hence, as $N \to \infty$, $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ have a non-zero asymptotic bias given by,

$$Bias_\gamma = G - \gamma; \quad Bias_{\sigma^2} = S - \sigma^2. \tag{33}$$

The explicit expressions of these asymptotic biases are given below in (34) and (35).

Proof . Since the SOU-process Y^ε is here a fixed stationary Gaussian process from which we sub-sample the observations U_n^ε with a fixed time-step Δ, the proof applies exactly the same generic principles as the proof of proposition (2) above, and we may directly apply the results from section (4) to the covariance estimators $\hat{r}_k^\epsilon$ for $k = 0, 1$ and to $\hat{\gamma}_\varepsilon = g\left(\hat{r}_0^\epsilon, \hat{r}_1^\epsilon\right)$ and $\hat{\sigma}_\varepsilon^2 = s\left(\hat{r}_0^\epsilon, \hat{r}_1^\epsilon\right)$, given by (31). □

As expected, *indirect estimation* of the OU process is less favorable than estimation based on *direct OU observations*, so that the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ are not consistent as $N \to \infty$, for a fixed value of ε and Δ. Instead, these estimators have non zero asymptotic biases $(G - \gamma)$ and $(S - \sigma^2)$ given by (32),(33), that are functions of Δ, ε.

The asymptotic biases do not remain bounded for all values of $\varepsilon \to 0$, $\Delta \to 0$. In the following proposition we derive the exact regime where it is possible to achieve asymptotic consistency of the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ in the limit of $\varepsilon \to 0$, $\Delta \to 0$.

Proposition 8 (Favorable Regime for Consistency) *As seen in proposition (7), for fixed ε and Δ, the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ both have non-zero asymptotic biases $Bias_\gamma$ and $Bias_{\sigma^2}$ as $N \to \infty$, which depend only on $\varepsilon, \Delta, \gamma, \sigma$. Assume now that $\varepsilon \to 0$, and for each ε select a number $N = N(\varepsilon)$ of indirect observations of Y_t^ε and a sub-sampling rate $\Delta = \Delta(\varepsilon)$ such that $\Delta \to 0$ and $N\Delta \to \infty$.*

Then, as $\varepsilon \to 0$, $Bias_\gamma$ and $Bias_{\sigma^2}$ tend to 0 if and only if $(\Delta/\varepsilon) \to \infty$.

Proof . From formula (28), we see that as $\varepsilon \to 0$, we have $K^\varepsilon(0) \to \sigma^2/2\gamma$; then the expression of S given by (32) shows that whenever $G \to \gamma$ as $\varepsilon \to 0$, we must also have $S \to \sigma^2$. Hence we only need to study the asymptotic behavior of $Bias_\gamma$. Note first that whenever $\Delta = \Delta(\varepsilon) \geq \varepsilon$ and $\varepsilon \to 0$, we have in view of (29) and (33),

$$Bias_\gamma = -(1/\Delta) \log\left(\frac{e^{-\gamma\varepsilon} + e^{\gamma\varepsilon} - 2}{2\left(e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon\right)}\right) \approx -\frac{\gamma\varepsilon}{3\Delta} \tag{34}$$

We have several cases to consider.

- **Case (a) :** Assume that $(\Delta/\varepsilon) \to \infty$ as $\varepsilon \to 0$. Then, for ε small enough, we have $\Delta \geq \varepsilon$, and (34) proves that $Bias_\gamma \to 0$, as $\varepsilon \to 0$, and, hence, $Bias_{\sigma^2} \to 0$.
- **Case (b) :** Let Δ/ε remain bounded as $\varepsilon \to 0$. Then, there exist a subsequence $\varepsilon \to 0$ such that $\Delta/\varepsilon \to L$ for some non negative L.
  - **Case (ba) :** If $L \geq 1$ then for ε small enough we have $\Delta \geq \varepsilon$ and hence, in view of (34), we have $Bias_\gamma \approx (-\gamma\varepsilon/(3\Delta))$ so that $Bias_\gamma$ tends to the nonzero limit $(-\gamma/(3L))$.
  - **Case (bb) :** Assume that $L < 1$. Then for ε small enough we have $\Delta < \varepsilon$ and hence, in view of (30) and (33), we have

$$Bias_\gamma = -(1/\Delta) \log\left(\frac{2\gamma(\varepsilon - \Delta)e^{\gamma\Delta} + e^{-\gamma\varepsilon}(e^{2\gamma\Delta} + 1) - 2}{2\left(e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon\right)}\right). \tag{35}$$

Taylor expansions with respect to ε in (35) easily shows that $Bias_\gamma$ tend to $(\gamma/L)\left(1/3 - L + L^2 + 2L^3/3\right)$, when $\varepsilon \to 0$. But the polynomial $\left(1/3 - L + L^2 + 2L^3/3\right)$ remains strictly positive for $0 \le L < 1$. Hence, $Bias_\gamma$ tends to a nonzero limit in case (bb).

□

Proposition (8), clearly defines a **favorable regime for adaptive sub-sampling**. We have seen that the asymptotic biases of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$, namely, $Bias_\gamma$ and $Bias_{\sigma^2}$, tend to 0 as $\varepsilon \to 0$ if and only if $(\Delta/\varepsilon) \to \infty$. This strongly indicates that optimal *adaptive sub-sampling* schemes from indirect observations based on Y^ε should provide $N = N(\varepsilon)$ observations $U_n^\varepsilon = Y_{n\Delta}^\varepsilon$ sub-sampled from Y_t^ε at time interval $\Delta = \Delta(\varepsilon)$, under the following set of simultaneous conditions,

$$\varepsilon \to 0; \quad \Delta \to 0; \quad \Delta/\varepsilon \to \infty; \quad N\Delta \to \infty. \tag{36}$$

These results highlight the necessity, as $\varepsilon \to 0$, to sub-sample the approximating process Y^ε with a *vanishing but coarse time-step* $\Delta(\varepsilon) >> \varepsilon$ to hope to obtain asymptotically consistent estimates of the underlying parameters.

Under *fixed rate sub-sampling*, applying the general results on the asymptotic properties of empirical covariance estimators based on the observations from a stationary Gaussian processes as described in [2, Chapter X], the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are asymptotically Gaussian, i.e., the random vector $\sqrt{N}\left(\hat{\gamma}_\varepsilon - G, \hat{\sigma}^2_\varepsilon - S\right)$ converges to a Gaussian distribution with mean zero, and covariance matrix dependent on the true parameters γ, σ^2, ε, and Δ.

Since, in particular, for each fixed ε, $\Delta > 0$, as $N \to \infty$, the empirical covariance estimators $\hat{r}_0^\varepsilon$, $\hat{r}_1^\varepsilon$ are asymptotically Gaussian [2]. The estimators, using (31), are given by $\hat{\gamma}_\varepsilon = g\left(\hat{r}_0^\varepsilon, \hat{r}_1^\varepsilon\right)$ and $\hat{\sigma}^2_\varepsilon = s\left(\hat{r}_0^\varepsilon, \hat{r}_1^\varepsilon\right)$, such that g, s have continuous second-order partial derivatives in a neighborhood of the true values r_0, r_1. Therefore, as $N \to \infty$, for each fixed ε, $\Delta > 0$, the estimators $\hat{\gamma}_\varepsilon = g\left(\hat{r}_0^\varepsilon, \hat{r}_1^\varepsilon\right)$ and $\hat{\sigma}^2_\varepsilon = s\left(\hat{r}_0^\varepsilon, \hat{r}_1^\varepsilon\right)$ are asymptotically Gaussian [2].

We now study these estimators under the conditions (36) in detail.

8 Adaptive Sub-sampling for Indirect Estimation

Proposition 9 (Asymptotics of the Covariances) *Consider an adaptive sub-sampling scheme, based on $N = N(\varepsilon)$ indirect observations extracted from Y_t^ε by sub-sampling with time steps $\Delta = \Delta(\varepsilon)$. Then, under the conditions* (36)*, the L_2 norms of the variables $(\hat{r}_0^\epsilon - r_0^\varepsilon)$, and $(\hat{r}_1^\epsilon - r_1^\varepsilon)$ converge to 0 with speeds of convergence proportional to $1/\sqrt{N\Delta}$.*

Moreover, for each $k = 0, 1$, the L_2 norm of $\sqrt{N\Delta}\,(\hat{r}_k^\epsilon - r_k^\varepsilon)$ converges to $(\sigma^2/(\gamma\sqrt{2\gamma}))$, which is identical to the asymptotic limit obtained when direct observations of the underlying OU process are available.

Proof . Define $J_k^\epsilon = E\left[\left(\hat{r}_k^\epsilon - r_k^\varepsilon\right)^2\right]$ for $k = 0, 1$ computed explicitly for $\Delta > \varepsilon$, by using proposition (1). Let

$$C_0 = \frac{\sigma^2(e^{-\gamma\varepsilon} + \gamma\varepsilon - 1)}{\gamma^3\varepsilon^2}, \quad C_1 = \frac{\sigma^2(e^{\gamma\varepsilon} + e^{-\gamma\varepsilon} - 2)}{2\gamma^3\varepsilon^2}, \quad \text{and} \quad b = e^{-\gamma\Delta},$$

then we have

$$J_0^\epsilon = \frac{2C_1^2}{N}\left(\frac{C_0^2}{C_1^2} + \frac{2b^2}{1-b^2} - \frac{2b^2\left(1-b^{2N}\right)}{N(1-b^2)^2}\right),$$

$$J_1^\epsilon = \frac{C_1^2}{N}\left(\frac{C_0^2}{C_1^2} + \frac{2C_0 b^2}{C_1} + \frac{b^2(3+b^2)}{1-b^2} - \frac{2b^2 B_1}{N}\right), \tag{37}$$

where $B_1 = \left[(C_0/C_1) + \left(1+2b^2-b^4-2b^{2N}\right)/\left(1-b^2\right)^2\right]$. From (37) we obtain bounds for J_0^ϵ and J_1^ϵ given by

$$J_0^\epsilon \le \frac{2C_0^2}{N} + \frac{2C_1^2}{\gamma N\Delta}, \qquad \text{and} \qquad J_1^\epsilon \le \frac{C_0^2}{N} + \frac{2C_0C_1e^{-\gamma\Delta}}{N} + \frac{2C_1^2}{\gamma N\Delta}.$$

These inequalities show that $J_k^\epsilon \to 0$ under the *adaptive sub-sampling scheme* defined in (36). The exact expressions in (37) gives, as $\varepsilon \to 0$,

$$(N\Delta)J_k^\epsilon \to \frac{\sigma^4}{2\gamma^3},$$

Therefore, the L_2-speeds of convergence for the empirical covariance estimators, as $\varepsilon \to 0$, are proportional to $1/\sqrt{N\Delta}$. □

Proposition 10 *For each N, Δ, ε, the random variables $Z_0 = Z_0(N,\Delta,\varepsilon)$ and $Z_1 = Z_1(N,\Delta,\varepsilon)$ defined by,*

$$Z_0 = \frac{(\hat{r}_0^\epsilon - r_0^\varepsilon)}{\sqrt{J_0^\epsilon}}, \qquad Z_1 = \frac{(\hat{r}_1^\epsilon - r_1^\varepsilon)}{\sqrt{J_1^\epsilon}}, \tag{38}$$

have mean 0, variance 1, and covariance $E[Z_0Z_1] = J_{01}^\epsilon/\sqrt{J_0^\epsilon J_1^\epsilon}$, where $J_0^\epsilon, J_1^\epsilon, J_{01}^\epsilon$ are given by (37), (40).

Then, under the conditions,

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad \Delta > \varepsilon,$$

the following first-order L_2 approximations for the empirical covariances $\hat{r}_k^\epsilon$ hold, namely,

$$\hat{r}_0^\epsilon = r_0^\varepsilon + \frac{\sqrt{2}r_0}{\sqrt{\gamma N\Delta}}Z_0 + \frac{Z_0}{\sqrt{N\Delta}}\left(O(\varepsilon^2) + O(\Delta^2) + O\left(\frac{1}{N\Delta}\right)\right),$$

$$\hat{r}_1^\epsilon = r_1^\varepsilon(\Delta) + \frac{\sqrt{2}r_0}{\sqrt{\gamma N\Delta}}Z_1 + \frac{Z_1}{\sqrt{N\Delta}}\left(O(\varepsilon^2) + O(\Delta^2) + O\left(\frac{1}{N\Delta}\right)\right), \tag{39}$$

where $O(h)$ is a deterministic function of h, bounded by a constant multiple of h.

Proof . Let $b = e^{-\gamma\Delta}$ and,

$$C_0 = \frac{\sigma^2(e^{-\gamma\varepsilon} + \gamma\varepsilon - 1)}{\gamma^3\varepsilon^2}, \quad C_1 = \frac{\sigma^2(e^{\gamma\varepsilon} + e^{-\gamma\varepsilon} - 2)}{2\gamma^3\varepsilon^2},$$

then, the exact expression for the covariance $J^{\epsilon}_{01} = E\left[\left(\hat{r}^{\epsilon}_0 - r^{\varepsilon}_0\right)\left(\hat{r}^{\epsilon}_1 - r^{\varepsilon}_1\right)\right]$, is given by

$$J^{\epsilon}_{01} = \frac{2C_1^2}{N}\left(\frac{2C_0 b}{C_1} + \frac{2b^3}{1-b^2} - \frac{B_2}{N}\right), \tag{40}$$

where

$$B_2 = \left(\frac{C_0 b}{C_1}\right) + \left(\frac{3b^3 - b^5 - b^{2N+1}(1+b^2)}{(1-b^2)^2}\right).$$

Using Taylor's expansions we obtain the following approximations,

$$J^{\epsilon}_0 = \frac{2r_0^2}{\gamma N\Delta}\left(1 + \frac{\gamma^2\Delta^2}{3} - \frac{1}{2\gamma N\Delta} + O(\Delta^4) + \frac{O(\Delta^2)}{N\Delta} + O(\varepsilon^2)\right),$$

$$J^{\epsilon}_1 = \frac{2r_0^2}{\gamma N\Delta}\left(1 - \frac{2\gamma^2\Delta^2}{3} + \gamma^3\Delta^3 - \frac{1}{2\gamma N\Delta} + O(\Delta^4) + \frac{O(\Delta^2)}{N\Delta} + O(\varepsilon^2)\right),$$

$$J^{\epsilon}_{01} = \frac{2r_0^2}{\gamma N\Delta}\left(1 - \frac{\gamma^2\Delta^2}{6} - \frac{1}{2\gamma N\Delta} + O(\Delta^4) + \frac{O(\Delta)}{N\Delta} + O(\varepsilon^2)\right),$$

from which we can deduce (39). □

The following theorem presents the key results of our study.

Theorem 1 (Consistency of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$) *Consider an adaptive sub-sampling scheme, based on $N = N(\varepsilon)$ indirect observations extracted from Y^{ε}_t by sub-sampling with time steps $\Delta = \Delta(\varepsilon)$. Let the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ of γ, σ^2, be given by* (31)*. Then, under the following conditions,*

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad \Delta/\varepsilon \to \infty, \tag{41}$$

the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ are asymptotically consistent, i.e., $\hat{\gamma}_\varepsilon \to \gamma$, $\hat{\sigma}^2_\varepsilon \to \sigma^2$ in L_2.

Moreover, the expected L_2-speed of convergence, proportional to $1/\sqrt{N\Delta}$, is achievable under the following conditions which are stronger than (41)*,*

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad N\varepsilon^2/\Delta < cte. \tag{42}$$

In particular, under stronger conditions than (42)*,* (41)*, namely,*

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta^2 \to \infty, \quad N\varepsilon^2/\Delta^2 \to 0, \tag{43}$$

the estimators are asymptotically efficient, and the asymptotic limit of the L_2-norms of the random variables $\sqrt{N\Delta}\,(\hat{\gamma}_\varepsilon - \gamma)$, $\sqrt{N}\left(\hat{\sigma}^2_\varepsilon - \sigma^2\right)$ converge, respectively, to $\sqrt{2\gamma}$, $\sigma^2\sqrt{2}$, exactly as in the case of direct observations.

Proof . Substitute the expressions for empirical covariance estimators $\hat{r}_k^\epsilon$, given by (38), in the expressions for the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ defined in (31). In particular, the drift estimator $\hat{\gamma}_\varepsilon$ is given by,

$$\hat{\gamma}_\varepsilon = \frac{-1}{\Delta} \log \left(\frac{e^{-\gamma\Delta} C_1 + \sqrt{J_1^\epsilon} Z_1}{C_0 + \sqrt{J_0^\epsilon} Z_0} \right).$$

Then, using Taylor expansions as $\varepsilon \to 0$ and using arguments similar to those given in the proof of proposition (6), we obtain the following first-order L_2-approximation for $\hat{\gamma}_\varepsilon$ given by,

$$\hat{\gamma}_\varepsilon = \gamma - \frac{\gamma\varepsilon}{3\Delta} - Z_\gamma + R_\gamma \times O\left(\frac{1}{N\Delta}\right) + \frac{\varepsilon}{\Delta} \times O(\varepsilon), \tag{44}$$

where the zero mean random variable Z_γ is given by,

$$Z_\gamma = \frac{e^{\gamma\Delta}\sqrt{J_1^\epsilon} Z_1}{\Delta C_1} - \frac{\sqrt{J_0^\epsilon} Z_0}{\Delta C_0}.$$

The L_2 norm of the random variable Z_γ using Taylor's expansion for $\varepsilon \to 0$, $\Delta \to 0$, $N\Delta \to \infty$ is approximated by,

$$\|Z_\gamma\|_{L^2}^2 = \frac{2\gamma}{N\Delta}\left(1 + 3\gamma\Delta - \frac{1+O(\Delta)}{2\gamma N\Delta} + O\left(\frac{\varepsilon}{\Delta}\right) + O(\Delta^2) + O(\varepsilon)\right). \tag{45}$$

The remainder term $R_\gamma = R_\gamma(\Delta, \varepsilon, N)$ is uniformly bounded in L_2 norm. Therefore, using (44) and (45), under the conditions (41) the estimator $\hat{\gamma}_\varepsilon$ converges in L_2 to γ.

To compute the L^2-speed of convergence we study

$$\sqrt{N\Delta}\,(\hat{\gamma}_\varepsilon - \gamma) = -\sqrt{N\Delta} Z_\gamma - \frac{\gamma\varepsilon\sqrt{N}}{3\sqrt{\Delta}} + R_\gamma \times O\left(\frac{1}{\sqrt{N\Delta}}\right). \tag{46}$$

Using (45), (46) we see that the L^2-norm of $\sqrt{N\Delta}\,(\hat{\gamma}_\varepsilon - \gamma)$ converges to a constant under conditions (42). Under the *adaptive sub-sampling scheme* (42), we assume $N\varepsilon^2/\Delta \to 0$ to deduce that the asymptotic variance of estimation errors converge to the same constant as in the case of *direct estimation* (see proposition (6)), i.e.,

$$\|\sqrt{N\Delta}\,(\hat{\gamma}_\varepsilon - \gamma)\,\|_{L^2}^2 \to 2\gamma.$$

Similarly, given the conditions (41), the diffusion estimator $\hat{\sigma}_\varepsilon^2 = 2\hat{\gamma}_\varepsilon \hat{r}_0^\epsilon$ converges in L_2 to the true value σ^2, and, hence, is asymptotically consistent. Furthermore, under conditions (43) we obtain,

$$\|\sqrt{N}\left(\hat{\sigma}_\varepsilon^2 - \sigma^2\right)\|_{L^2}^2 \to 2\sigma^4.$$

□

The main conclusion of theorem (1) is that under conditions (42) the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$, based on indirect estimation, are *asymptotically consistent* estimators of γ, σ^2, with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$.

A natural objective is to optimally select $\Delta = \Delta(\varepsilon)$ and $N = N(\varepsilon)$, verifying the conditions (42), in order to achieve the fastest speed of convergence. A *pragmatic interpretation* of the conditions (42) is that, as $\varepsilon \to 0$, one selects $\Delta = \Delta(\varepsilon)$ such that

$$\Delta \to 0, \quad \Delta >> \varepsilon, \quad \text{and N verifies, } (1/\Delta) << N < cte(\Delta/\varepsilon^2). \tag{47}$$

The L_2-speed of convergence $(1/\sqrt{N\Delta})$ of our estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ then verifies,

$$cte\left(\frac{\varepsilon}{\Delta}\right) < \frac{1}{\sqrt{N\Delta}} << 1. \tag{48}$$

Clearly, the lower bound (ε/Δ) in (48) is the best L_2-speed of convergence achievable under the conditions (42). This speed is attained when $N \sim \Delta/\varepsilon^2 \to \infty$, which corresponds to a global time interval of observations $T^* = N\Delta = cte(\Delta^2/\varepsilon^2)$.

Choosing a global time interval of observations $T >> T^* \to \infty$ will not improve the accuracy, since, the L_2 errors will then be dominated by $(\varepsilon/\Delta) >> (1/\sqrt{N\Delta})$. This, indeed, provides evidence that under indirect estimation, observing the data on an increasing time interval $N\Delta$ will not improve by itself the accuracy of the estimators, and coarse graining (*i.e.*, $\Delta >> \varepsilon$) of the data is necessary to reduce the estimation errors.

In the following corollary we provide a particular example of the optimal criterion identified by the pragmatic interpretation (47).

Corollary 1 (Power Law Criterion for Optimal Sub-Sampling)
As $\varepsilon \to 0$, assume that $N(\varepsilon)$ and $\Delta(\varepsilon)$ are given by powers of ε, namely, $N(\varepsilon) = \varepsilon^{-\eta}$, $\Delta(\varepsilon) = \varepsilon^{\alpha}$. Then,

1. *as $\varepsilon \to 0$, for any α, η such that $\alpha \in (0,1)$, $\eta > \alpha$, the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are asymptotically consistent in L_2 norm.*
2. *Moreover, as $\varepsilon \to 0$, under stronger conditions, namely, for any α, η such that $\alpha \in (0,1)$, $\alpha < \eta \leq 2-\alpha$, the estimators converge with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta} = \varepsilon^{(\eta-\alpha)/2}$.*
3. *The best speed of convergence are reached when $\alpha > 0$ is close to 0, and $\eta = 2-\alpha$. Then, we obtain $\Delta = \varepsilon^{\alpha}$, $N = \varepsilon^{-(2-\alpha)}$, and the global time of observations $N\Delta = \varepsilon^{-2(1-\alpha)}$.*

9 Numerical Simulations

We now study numerically a few typical examples of adaptive sub-sampling schemes ensuring asymptotic consistency of estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$. In view of the corollary (1), we let $\Delta(\varepsilon) = \varepsilon^{\alpha}$ where $\alpha \in (0,1)$, and the number of observations $N >> (\Delta/\varepsilon^2)$. The following numerical results show that as $\varepsilon \to 0$, $Bias_\gamma(\Delta,\varepsilon)$ and $Bias_{\sigma^2}(\Delta,\varepsilon)$ converge to 0 if and only if $(\Delta(\varepsilon)/\varepsilon) \to \infty$ (See proposition (8)). As evident in the following numerical study and from corollary (1) for smaller values of $\alpha \in (0,1)$, the convergence of the $Bias_\gamma$ and $Bias_{\sigma^2}$ to zero is faster.

We generate numerical discrete simulations for the trajectory X_t of the OU-process with fixed parameters $\gamma = 3.2625$ and $\sigma = 6.7500$. Each associated SOU-process trajectory Y_t^ε is computed by direct integration of the discretized trajectory X_t on a sliding time window of duration ε . The N observed data are then obtained by sub-sampling the discretized SOU-trajectory Y_t^ε with step size Δ. The goal was to verify the analytical results derived above on indirect sub-sampling estimation of the underlying parameters.

The underlying discretized trajectory of X_t is generated using a hybrid of Euler-Maruyama and second-order Runge-Kutta discretization schemes for the SDE (1), with a time-step length of $d = 10^{-4}$ and total time interval $T = 900$, thus providing 9×10^6 points of OU-trajectory. To generate SOU-observations, we average the simulated OU observations over a sliding window of length ε, for the following values of ε,

$$\varepsilon \quad = \quad 0.01, \quad 0.05, \quad 0.1, \quad 0.15, \quad 0.2, \quad 0.25, \quad 0.3.$$

We consider 3 examples of *adaptive sub-sampling* schemes, namely, when observations are sub-sampled with time-step $\Delta(\varepsilon) = \varepsilon^{0.5}$, $\Delta(\varepsilon) = \varepsilon$, and $\Delta(\varepsilon) = \varepsilon^2$. In each one of these 3 cases, for each simulated trajectory of the SOU process, we compute the subsampled estimators $\hat{\gamma}_N$ and $\hat{\sigma}_N^2$ given by (31). Figure (1) shows numerical verification of the consistency results obtained in section (6). Errors (in %) in the figure is defined to be the absolute value of the relative bias in the estimates. For instance, for the error in the estimation of γ, we have

$$Error = \left| \frac{\hat{\gamma}_N - \gamma}{\gamma} \right|.$$

1. **Case** $\Delta(\varepsilon) = \varepsilon^{0.5}$: Results are displayed in the top part of Figure (1). The empirical relative bias (errors) of sub-sampled estimators tend to zero as $\varepsilon \to 0$, as expected, since $\Delta(\varepsilon)/\varepsilon \to \infty$ in this case.
2. **Case** $\Delta(\varepsilon) = \varepsilon$: Results are displayed in the middle part of Figure (1). The empirical relative bias (errors) of the sub-sampled estimators converge to a non zero value, as $\varepsilon \to 0$, as expected, since $\Delta(\varepsilon)/\varepsilon$, is bounded in this case.
   Formula (34) for the asymptotic bias give $Bias_\gamma \approx -\gamma/3$ and $Bias_{\sigma^2} \approx -\sigma^2/3$, which fit very well with the numerical results.
3. **Case** $\Delta(\varepsilon) = \varepsilon^2$: Results are displayed in the bottom part of Figure (1). The empirical relative bias (errors) of the sub-sampled estimators increase as $\varepsilon \to 0$, as expected, since $\Delta(\varepsilon)/\varepsilon \to 0$ in this case.

10 A practical example : the Additive Triad

We outline here a concrete example encountered in simplified dynamic models of atmospheric evolution, when only 3 main modes are kept in the Additive Triad model [13]. This example will be studied in detail elsewhere, and is only sketched here. We provide results obtained from estimating parameters

γ and σ^2 in SDE (1) when data is sub-sampled from the slow mode in the Additive Triad model.

The additive triad model comprises of the stochastic process $[x_t, y_t, z_t]$ in R^3, where the slow mode x_t and the two fast modes y_t, z_t are driven by the equations,

$$\begin{aligned} dx_t &= A_1 y_t z_t \frac{dt}{\varepsilon}, \\ dy_t &= A_2 x_t z_t \frac{dt}{\varepsilon} - g_1 y_t \frac{dt}{\varepsilon^2} + s_1 \frac{dW_1(t)}{\varepsilon}, \\ dz_t &= A_3 x_t y_t \frac{dt}{\varepsilon} - g_2 z_t \frac{dt}{\varepsilon^2} + s_2 \frac{dW_2(t)}{\varepsilon}, \end{aligned} \tag{49}$$

where $A_1 + A_2 + A_3 = 0$, g_i, s_i are strictly positive and $\varepsilon > 0$ is the scale separation parameter, and where W_1, W_2 are Brownian motions. It is well known [13] that in the limit of infinite scale separation as $\varepsilon \to 0$, the slow mode x_t converges weakly to the OU process X_t with parameters γ and σ given by

$$\gamma = \frac{-A_1}{2(g_1 + g_2)} \left(\frac{A_2 s_2^2}{g_2} + \frac{A_3 s_1^2}{g_1} \right), \qquad \sigma^2 = \frac{(A_1 s_1 s_2)^2}{2 g_1 g_2 (g_1 + g_2)}. \tag{50}$$

In this context estimations of the parameters γ and σ must be performed using only indirect observations of X_t generated by the slow mode x_t of the Additive Triad model. This example is quite close to the SOU process analyzed above and we have carried out numerical simulations of the additive triad model using the following parameter values,

$$A_1 = 0.9, \quad A_2 = -0.4, \quad A_3 = -0.5, \quad g_1 = 1, \quad s_1 = 3, \quad g_2 = 1, \quad s_2 = 5.$$

The associated reduced model parameters are, $\gamma = 3.2625$, $\sigma = 6.7500$. We consider the estimators $\hat{\gamma}_N$ and σ_N^2, given by (13), based on $(N+1)$ observations sub-sampled from the slow mode x such that $U_n = x_{n\Delta}$. We consider three adaptive sub-sampling strategies for the indirect estimation of the OU SDE from the data generated by the additive triad model in (49).

Top part of the Figure (2) demonstrates that when $\left(\Delta(\varepsilon)/\varepsilon^2\right) \to \infty$ estimates for γ and σ^2 are consistent with respect to the theoretical results in (50). On the other hand, errors remain bounded away from zero for the adaptive sub-sampling strategy such that $\left(\Delta(\varepsilon)/\varepsilon^2\right)$ is bounded. This is depicted in the middle part of Figure (2) where $\left(\Delta(\varepsilon)/\varepsilon^2\right)$ tends to a non-zero value, and hence, the errors converge to a constant strictly greater than zero. The bottom part of Figure (2) is based on sub-sampling scheme such that $\left(\Delta(\varepsilon)/\varepsilon^2\right) \to 0$, and the corresponding estimation errors increase to 100%. Therefore, sub-sampling strategy $\Delta >> \varepsilon^2$ is the favorable sub-sampling regime for the estimation of the OU SDE from the triad data. The nature of the results is similar to the ones obtained for the SOU process.

Effectively, here ε^2 plays the same role as ε in the SOU process (c.f. Figures 1 and 2). This can be understood by analyzing the correlation function of x_t for small lags. It can be shown that the correlation function of x_t scales as $1 - cte(\tau^2\varepsilon^{-2})$, where τ is the lag and ε is the parameter in (49). On the other hand, the correlation function of Y_t^ε scales as $1 - cte(\tau^2\varepsilon^{-1})$. Therefore, $\Delta >> \varepsilon^2$ is the correct sub-sampling criteria in the triad model in (49), and the analogous *adaptive sub-sampling scheme* to ensure consistency of estimators is given by the following conditions,

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad (N\varepsilon^4/\Delta) < cte.$$

11 Conclusions and further research

The main result of our study is the characterization of *optimal adaptive sub-sampling schemes* in a Gaussian context. The goal was to consistently estimate the drift and diffusion parameters γ and σ of a non observable OU-process X_t, using $N(\varepsilon)$ observations extracted by sub-sampling, at time intervals $\Delta(\varepsilon)$, an approximating process Y_t^ε which tends to X_t in L_2 as $\varepsilon \to 0$. We obtain explicit asymptotic results for the estimation errors, and derive sufficient conditions on $N(\varepsilon)$, $\Delta(\varepsilon)$ ensuring that $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ are asymptotically consistent estimators of the unknown parameters γ, σ^2 of the *unobserved* OU process. We also analyze the speed of convergence of consistent estimators, and show that under explicit stronger conditions, as $\varepsilon \to 0$, the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$, have an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$.

Our paper focuses on the favorable situation where the OU process X_t is approximated by the Gaussian SOU processes Y_t^ε, to characterize explicitly the family of optimal sub-sampling regimes leading to consistent estimators having the best L_2-speeds of convergence. This specific framework replicates the scenario observed in several applications where a mismatch between the data and the stochastic model impedes the estimation procedure, so that appropriate adaptive sub-sampling schemes become necessary to obtain consistent estimates and good L2-speeds of convergence.

In an ongoing study, we will extend the main results of this paper to a much wider class of stationary Gaussian processes X_t, and to arbitrary non-Gaussian stationary processes Y_t^ε such that as $\varepsilon \to 0$, the random variables Y_t^ε tend to X_t in L_p for some $p > 2$. In such generic cases, the adequate convergence speed of good adaptive sub-sampling schemes can be identified by expanding the correlation function of Y^ε for small lags, or alternatively by the L_p-speed of convergence of the approximating process Y_t^ε to X_t, as $\varepsilon \to 0$. In this paper, we have briefly illustrated the use of a small lag expansion of correlation functions, while comparing the SOU process to the Additive triad model.

From a pragmatic point of view, insufficient sub-sampling can lead to large errors in practical fitting of stochastic models to physical data intensively sampled from complex dynamic systems. On the other hand, optimal regimes for efficient sub-sampling depend on an indexation parameter ε which is rarely known "intrinsically", or even explicitly. Therefore, development of efficient sub-sampling tests based on discrete datasets alone are necessary. One of the natural approaches we are exploring in this direction is the empirical robustness of estimators with respect to multiple sub-sampling of large finite data sets with different values of Δ. This is equivalent to treating estimators as functions of the sub-sampling time step and analyzing their behavior as Δ decreases. This points out the part played by the data points between $X(t)$ and $X(t+\Delta)$, which have a potential efficient impact to determine concretely wether a given Δ defines an empirically adequate sub-sampling rate. In this context, understanding how sub-sampling affects the bias terms is the key for constructing accurate and efficient estimators using only approximate data.

12 Appendix: Moments of SOU process

We outline briefly the explicit computation of moments in (27) for the SOU-process Y_t^ε given by (26). The mean of Y_t^ε is obviously 0, and the covariance function of Y_t^ε at time lag h is given by,

$$K^\varepsilon(h) = \frac{1}{\varepsilon}\left(\int_{t+h-\varepsilon}^{t+h} E[X_s Y_t^\varepsilon] ds\right), \tag{51}$$

Using the stationary covariances of X_t, given by (4), for each $s \in [t+h-\varepsilon, t+h]$, where $h \geq \varepsilon$ we obtain,

$$E[X_s Y_t^\varepsilon] = \sigma^2 e^{-\gamma s}\left(e^{\gamma t} - e^{\gamma(t-\varepsilon)}\right)/(2\gamma^2\varepsilon). \tag{52}$$

Using (51) and (52) the covariance function $K^\varepsilon(h)$, for $h \geq \varepsilon$ is given by,

$$K^\varepsilon(h) = \left(\sigma^2 e^{-\gamma h}(e^{-\gamma\varepsilon} + e^{\gamma\varepsilon} - 2)\right)/(2\gamma^3\varepsilon^2).$$

For the case when $0 \leq h < \varepsilon$, using (51) we have

$$K^\varepsilon(h) = \frac{1}{\varepsilon}\left(\int_{t+h-\varepsilon}^{t} E[X_s Y_t^\varepsilon] ds + \int_t^{t+h} E[X_s Y_t^\varepsilon] ds\right),$$

where $E[X_s Y_t^\varepsilon]$ can be computed using the stationary covariances of X_t to give,

$$K^\varepsilon(h) = \sigma^2 e^{-\gamma h}\left\{2\gamma(\varepsilon - h)e^{\gamma\Delta} + e^{-\gamma\varepsilon}(e^{2\gamma h} + 1) - 2\right\}/(2\gamma^3\varepsilon^2).$$

Acknowledgments.

Research of I. Timofeyev and R. Azencott is partially supported by the NSF Grants DMS-0713793 and DMS-0811153, respectively.

References

1. Alperovich, T., Sopasakis, A.: Modeling highway traffic with stochastic dynamics. J.Stat.Phys **133**, 1083–1105 (2008)
2. Azencott, R., Dacunha-Castelle, D.: Series of Irregular Observations: Forecasting and Model Building. Springer (1986)
3. Chorin, A.J., Hald, O.H., Kupferman, R.: Optimal prediction and the Mori-Zwanzig representation of irreversible processes. Proc. Nat. Acad. Sci. USA **97**, 2968–2973 (2000)
4. Chorin, A.J., Hald, O.H., Kupferman, R.: Optimal prediction with memory. Physica D **166**, 239–257 (2002)
5. Chorin, A.J., Kast, A.P., Kupferman, R.: Unresolved computation and optimal prediction. Comm. Pure Appl. Math. **52**, 1231–1254 (1999)
6. Crommelin, D., Vanden-Eijnden, E.: Subgrid-scale parameterization with conditional Markov chains. J. Atmos. Sci. **65**, 2661–2675 (2008)
7. DelSole, T.: A fundamental limitation of Markov models. J. Atmos. Sci. **57**, 2158–2168 (2000)
8. Horenko, I., Dittmer, E., Fischer, A., Schutte, C.: Automated model reduction for complex systems exhibiting metastability. Multiscale Model. Simul. **5(3)**, 802–827 (2006)
9. Katsoulakis, M., Majda, A., Sopasakis, A.: Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 1, deterministic closures. Comm. Math. Sci. **2**, 255–294 (2004)
10. Katsoulakis, M., Majda, A., Sopasakis, A.: Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 2, stochastic closures. Comm. Math. Sci. **3**, 453–478 (2005)
11. Katsoulakis, M., Majda, A., Sopasakis, A.: Intermittency, matastability and coarse graining for coupled deterministic-stochastic lattice systems. Nonlinearity **19**(1–27) (2006)
12. Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: A mathematics framework for stochastic climate models. Comm. Pure Appl. Math. **54**, 891–974 (2001)
13. Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: A priori tests of a stochastic mode reduction strategy. Physica D **170**, 206–252 (2002)
14. Majda, A.J., Timofeyev, I., Vanden-Eijnden, E.: Stochastic models for selected slow variables in large deterministic systems. Nonlinearity **19(4)**, 769–794 (2006)
15. Papavasiliou, A., Pavliotis, G.A., Stuart, A.: Maximum likelihood drift estimation for multiscale diffusions. Stochastic Processes and their Applications **119**, 3173–3210 (2009)
16. Pavliotis, G.A., Stuart, A.: Parameter estimation for multiscale diffusions. J. Stat. Phys. **127**, 741–781 (2007)
17. Schutte, C., Walter, J., Hartmann, C., Huisinga, W.: An averaging principle for fast degrees of freedom exhibiting long-term correlations. Multiscale Model. Simul. **2(3)**, 501–526 (2004)

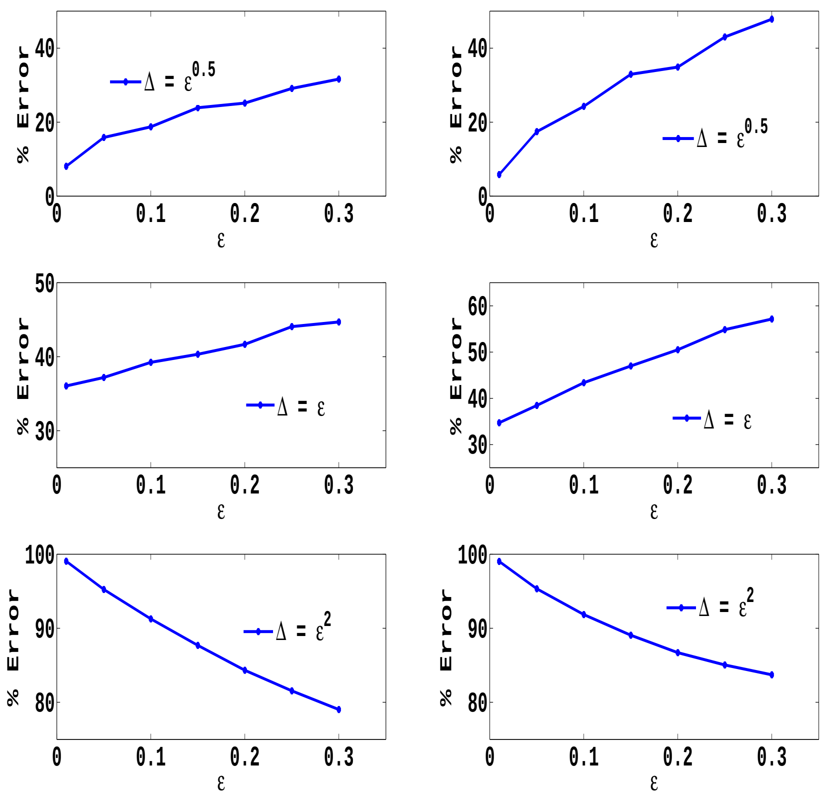

Fig. 1 Relative (%) Errors in MLEs for γ and σ^2 based on observations from the **SOU process** and sub-sampled with three different strategies. **Left part** - Relative errors for γ, **Right part** - Relative errors in σ^2. **Top part** - Sub-sampling with $\Delta = \varepsilon^{0.5}$: The errors converge to 0 with speed of convergence proportional to $\varepsilon^{0.5}$. **Middle part** - Sub-sampling with $\Delta = \varepsilon$: The errors converge to a constant ($\approx$ 33%) with speeds of convergence proportional to ε. **Bottom part** - Sub-sampling with $\Delta = \varepsilon^2$: The errors increase to 100%.

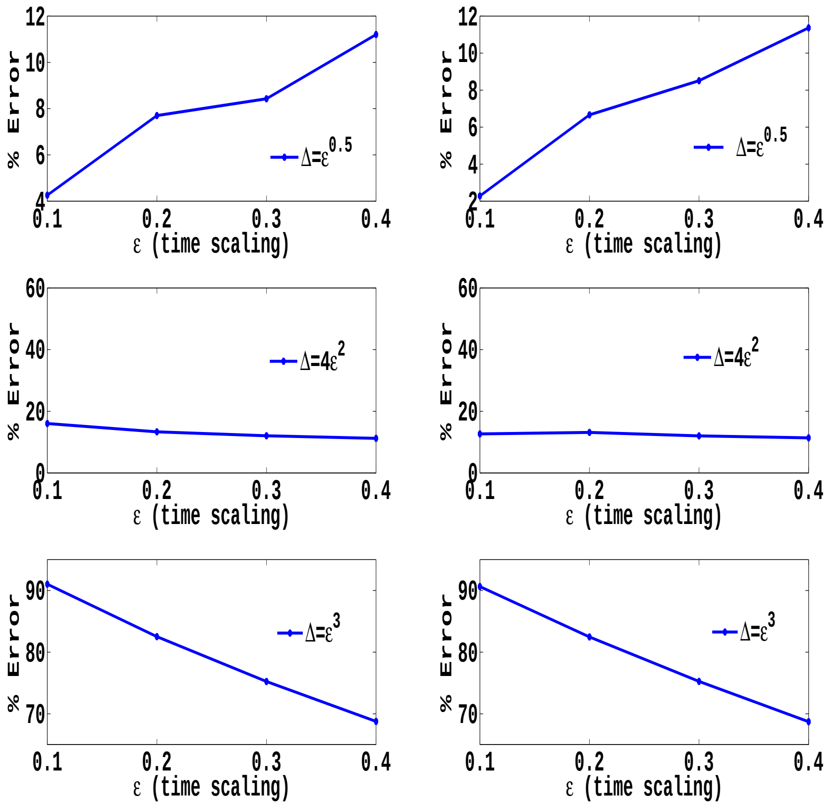

Fig. 2 Relative (%) Errors in MLEs for γ and σ^2 based on observations from the **Additive triad model** and sub-sampled with three different strategies. **Left part** - Relative errors for γ, **Right part** - Relative errors in σ^2. **Top part** - Sub-sampling with $\Delta = \varepsilon^{0.5}$: The errors converge to 0. **Middle part** - Sub-sampling with $\Delta = 4\varepsilon^2$: The errors converge to a constant. **Bottom part** - Sub-sampling with $\Delta = \varepsilon^3$: The errors increase to 100%.