SUB-SAMPLING AND PARAMETRIC ESTIMATION FOR MULTISCALE DYNAMICS *

R. AZENCOTT [†], A. BERI [‡], A. JAIN [§], AND I. TIMOFEYEV [¶]

Abstract. We study the problem of adequate data sub-sampling for consistent parametric estimation of unobservable stochastic differential equations (SDEs), when the data are generated by multiscale dynamic systems approximating these SDEs in some suitable sense. The challenge is that the approximation accuracy is scale dependent, and degrades at very small temporal scales. Therefore, maximum likelihood parametric estimation yields inconsistent results when the sub-sampling time-step is too small. We use data from three multiscale dynamic systems, the Additive Triad, the Truncated Burgers-Hopf models, and the Model with the Fast-Oscillating Potential to illustrate this sub-sampling problem. In addition, we also discuss an important practical question of constructing the bias-corrected estimators for a fixed but unknown value of the multiscale parameter.

Key words. Parametric Estimation, Stochastic Differential Equations, Sub-sampling

subject classifications. 62M05, 60J60, 60H10, 60F05

1. Introduction Stochastic modeling of discrete dynamic data has been an active area of research for many decades. This is motivated by the fact that in practice, observations are discrete and, typically, sampled with a fixed predefined time-step. Major examples include, for instance, stochastic modeling of multiscale financial data [3, 36, 12, 15], reduced stochastic description of the large-scale structures in the atmosphere-ocean (e.g. [22, 14, 7, 20]), and stochastic models for conformational changes of proteins [34, 16, 17]. In most situations, the exact dynamic model, even if it is known, is too complex and too costly to be used directly in numerical simulations. Thus, various stochastic modeling techniques have been designed to provide effective simplified descriptions of the process of interest. For instance, in the atmosphere/ocean problems one is interested in reproducing the statistical behavior of the leading empirical orthogonal functions. Therefore, typically, the discrete dynamic data are not generated by the proposed stochastic model; instead the goal of stochastic modeling is to recover long-term statistical properties of the data generated by the full model. Thus, the effective stochastic model is essentially used as as an approximation of the true dynamics in statistical sense.

The multiscale nature of complex dynamics has been an important research focus in this context. It has been recognized that many systems involve multiple interacting temporal and spatial scales, which cannot be disentangled in a trivial manner. Behavior of the atmosphere-ocean structures and protein folding fall in this category. In the last few years, several publications addressed the importance of multiscale effects in data-driven stochastic modeling [9, 10]. In particular, the sub-sampling issue has been addressed previously in the context of financial data [3, 36, 12], homogenized diffusions [32, 31], and non-parametric estimation [9].

---

*

[†]Department of Mathematics, University of Houston, Houston, TX (`razencot@math.uh.edu`); Emeritus Professor, Ecole Normale Superieure, Paris, France.

[‡]Mathematical Biosciences Institute, Ohio State University, Columbus, OH (`beri.3@mbi.osu.edu`).

[§]Department of Mathematics, University of Houston, Houston, TX (`ankita@math.uh.edu`) and Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, (`Ankita.Jain@nd.edu`).

[¶]Department of Mathematics, University of Houston, Houston, TX (`ilya@math.uh.edu`).

In two recent papers we have provided a rigorous foundation for the parametric estimation of Gaussian diffusions X_t from discrete approximate data [5, 6]. In particular, we considered the asymptotic behavior of the Maximum Likelihood Estimators for the unknown parameters of the diffusion X_t, when these estimators are computed from an observable multiscale approximating process Y_t^ϵ. The main emphasis was on the limiting behavior of these estimators as the multiscale parameter ϵ tends to zero. These two papers establish a connection between multiscale effects and the behavior of such Maximum Likelihood Estimators computed from discrete datasets sub-sampled from Y_t^ϵ. In particular, we demonstrated that for consistent estimation of the diffusion parameters the underlying dataset has to be sub-sampled with time-steps constrained by specific sub-sampling criteria, depending on the value of the multiscale parameter ϵ. Otherwise, if these sub-sampling criteria are violated, the estimated underlying diffusion model will not reproduce the statistical features of the data and the corresponding parameter estimators will be biased even in the limit $\epsilon \to 0$.

The main emphasis of the current paper is two-fold. First, extensive numerical investigation of our sub-sampling criteria in the context of homogenized models is performed. Second, we describe the construction of new bias-corrected estimators from datasets generated by an approximate multiscale model with an *unknown* fixed parameter ϵ. This is an important point, addressing a practical question of stochastic modeling of large-scale structures in multiscale high-dimensional systems. Typically, in such situations the time-series of the large-scales are available from numerical simulations of the full model, but the exact value of the multiscale parameter is unknown, since it does not enter the full equations explicitly.

To deduce the appropriate sub-sampling regime and construct the bias-corrected estimators we consider the Maximum Likelihood Estimators $\hat{\gamma}$ of the underlying diffusion parameter γ as functions of the sub-sampling time-step, $\hat{\gamma} \equiv \hat{\gamma}(\Delta)$. In this paper we concentrate on the Maximum Likelihood Estimators in the vanishing lags regime, i.e. computed for the small value of the sub-sampling step, Δ. The situation with non-vanishing lags is addressed in [6, 4], but this is not the subject of this paper. Since the Maximum Likelihood Estimators are expressed through the lagged covariance function (also called the correlation function), the behavior of the lagged covariance for small lags is crucial for understanding the behavior of the estimators. In this paper we consider the behavior of such parameter estimators when the discrete data are sub-sampled from trajectories generated by multiscale diffusions. We infer the behavior of the estimators as the sub-sampling step Δ is varied by comparing the asymptotic behavior of the correlation function in multiscale diffusions with the correlation function of the Smoothed Ornstein-Uhlenbeck process considered in [5]. Moreover, we also develop an approach for constructing the bias-corrected estimators when the data are generated from a trajectory $Y_{n\Delta}^\epsilon$ with a fixed, but unknown value of the multiscale parameter. The implementation of this approach and validating results are discussed in sections 2.2, 3.2, and 5.2. In particular, we show that this approach does not require the knowledge of the actual value of the scale-separation parameter ϵ, and enables us to construct an unbiased estimator by analysis of the curve $\hat{\gamma}(\Delta)$ vs Δ. In the examples considered here this curve is a straight line and we show that the slope estimator is an unbiased estimator for the damping parameter γ. In other examples the curve $\hat{\gamma}(\Delta)$ vs Δ can be more complicated, but we expect that in each particular case unbiased estimators can be constructed using nonlinear regression for the shape of this curve. Thus, we expect the approach developed in this paper to be

applicable to other systems, including nonlinear models and systems of SDEs. This will be discussed in subsequent papers.

The rest of the paper is organized as follows. In section 2 we discuss the indirect observability framework, outline the general form of the multiscale approximating processes considered in this paper, and recall briefly major results for the Ornstein-Uhlenbeck process from [5]. In sections 3 and 4 we consider two prototype multiscale approximating processes given by the triad model and fast-oscillating potential, respectively. In section 5 we introduce a more realistic dynamic model - the truncated Burgers-Hopf equation - and present results for parametric estimation of the modified truncated Burgers-Hopf model.

2. Indirect Observability Framework

The mathematical description of the indirect observability framework is a formal way to address an important case when the nature of the observed process is not known exactly or too complex to use in numerical/analytical calculations; instead, it is desirable to approximate this process by a suitable stochastic process X_t with matching statistical features. The observable process is denoted as Y_t^ϵ where ϵ is the scale-separation parameter in multiscale systems. The limit process of interest is denoted as X_t, but it is not observed directly; instead the parameters of the stochastic model for X_t are inferred from the data sub-sampled from Y_t^ϵ. Here Y_t^ϵ is also referred to as an approximating process in this paper.

To test the consistency of the estimation procedure we consider examples when Y_t^ϵ is such that $Y_t^\epsilon \to X_t$ as $\epsilon \to 0$ in some suitable sense. In the examples considered here the limiting behavior of Y_t^ϵ can be derived explicitly via a fairly classical homogenization procedures, which we only outline briefly. Nevertheless, the analytical expression for the limit stochastic model are only used to test the behavior of the Maximum Likelihood Estimators as $\epsilon \to 0$. In particular, since X_t is the limit process of Y_t^ϵ as $\epsilon \to 0$, it is desirable that the maximum likelihood parameter estimators computed from discrete data sub-sampled from Y_t^ϵ converge to the "true" parameter values as $\epsilon \to 0$. The homogenization procedure is used to derive the explicit equations for the limiting process and the "true" values of the parameters. We strongly emphasize that the homogenization procedure is only used here to test the consistency of parametric estimation; the maximum likelihood parameter estimators do not rely at all on the homogenization technique nor on the statistical knowledge of fast scales.

One practical example where the indirect observability framework is widely used is the data-driven stochastic modeling of slow variables in multiscale systems. In particular, there has been a significant effort to develop effective stochastic models for slow variables (often represented by leading empirical orthogonal functions) in geophysical applications (see e.g. [2, 11]). Such multiscale systems can be schematically represented as

$$dx_t = f(x_t)dt + \frac{1}{\epsilon}g(x_t, y_t)dt$$

$$(2.1)$$

$$dy_t = \frac{1}{\epsilon}h(x_t, y_t)dt + \frac{1}{\epsilon^2}q(y_t)dt + \frac{1}{\epsilon}s(y_t)dW_t,$$

where W_t is the Brownian motion and ϵ represents the scale separation parameter. A formal asymptotic theory as $\epsilon \to 0$ has been developed in [19, 18, 21, 30] (see also [25, 26, 27, 28, 14, 13] for more recent examples and applications) and it has been proved that there exists X_t such that $x_t \to X_t$ as $\epsilon \to 0$ in some suitable sense. There-

fore, in such multiscale systems, X_t plays the role of the unobserved limiting process and $Y_t^\epsilon \equiv x_t$ plays the role of the observed process. Often, a functional form of the effective stochastic model for X_t can be derived, but the model coefficients, typically, depend on the statistical behavior of the fast variables (y_t in the notation of (2.1)). Therefore, it is necessary to develop efficient parametric estimation techniques using only observations of the approximate process, Y_t^ϵ.

Moreover, often the exact value of the multiscale parameter ϵ is unknown in simulations of the full coupled system in (2.1). This is a typical situation in fluid dynamics where the multiscale parameter does not enter the partial differential equations describing the motion of the fluid explicitly. Instead, the multiscale nature manifests itself through the time-scale separation between different empirical orthogonal function (e.g. Fourier coefficients in simulations with periodic boundary conditions). Therefore, it is necessary to develop techniques for estimating the coefficients in the limiting process using the dataset generated from the simulations of multiscale systems with a fixed, but *unknown* value of ϵ.

Another multiscale example which is considered in this paper is the stochastic differential equation with a fast-oscillating potential

$$dx_t = -V'(x_t)dt - \frac{1}{\epsilon}p'(\frac{x_t}{\epsilon}) + sdW_t, \tag{2.2}$$

where $V(x_t)$ is the regular potential, $p(y)$ is the fast-oscillating potential, and W_t is the Brownian motion. Similar to the multiscale model presented earlier, equations for the limiting process can be derived explicitly in many cases. Here we study a particular case considered in [32] when the limiting process is the Ornstein-Uhlenbeck process. Estimating effective models with data generated from stochastic differential equations falls under the indirect estimation framework, since the limiting process is not observed directly.

Connection with previous work. In [5, 6] we discussed in detail the sub-sampling strategy under the L^p (with $p = 2$ or 4) convergence, i.e. $||Y_t^\epsilon - X_t||_p < \epsilon$ as $\epsilon \to 0$ for Gaussian limiting processes X_t. We considered the case when the vector of unknown parameters $\boldsymbol{\theta}$ in the SDE driving the limiting process X_t can be uniquely expressed as a differentiable function of lagged covariances $K(X_t, u_k) = \mathbb{E}[X_t X_{t+u_k}]$ with lags u_k, $k = 1, \ldots, J$. The behavior of the stationary lagged covariances $K(Y_t^\epsilon, u_k) = \mathbb{E}\left[Y_t^\epsilon Y_{t+u_k}^\epsilon\right]$ is crucial for developing the optimal sub-sampling strategy resulting in consistent estimation under indirect observability (when the covariances in the formula for parameters are estimated from the data of Y_t^ϵ instead of X_t). In particular, we proved that the speed of convergence of the lagged covariances $||K(Y_t^\epsilon, u_k) - K(X_t, u_k)||_2 \le \rho(\epsilon)$ determines the speed of convergence for the estimators $\hat{\boldsymbol{\theta}}$, which is given by $||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 \le C\left(\rho(\epsilon) + 1/\sqrt{T}\right)$, where $\rho(\epsilon)$ reflects the indirect observability of the process X_t and $1/\sqrt{T}$ is a typical term arising due to the finite observation time T. In previous work we considered the case when the speed of convergence of covariances is identical for different lags, which may not hold in general. Often some parameters may not depend on all covariances and, thus, the speed of convergence for estimators for each individual parameter may vary.

Parameters of the Ornstein-Uhlenbeck process can be uniquely expressed through the covariances $r_0 = \mathbb{E}[(X_t)^2]$ and $r_1 = \mathbb{E}[X_t X_{t+\Delta}]$ (see (2.5) in the next section). Therefore, Ornstein-Uhlenbeck process falls in the category of processes discussed above. In fact, the two quantities which are used for computing the drift and diffusion parameters are r_1/r_0 and r_0. Therefore, if the sub-sampling step Δ is kept

fixed, the speed of convergence for the drift estimator depends only on the behavior of $\hat{r}_1/\hat{r}_0$ and the speed of convergence for the diffusion estimator depends on both, $\hat{r}_1/\hat{r}_0$ and $\hat{r}_0$ (where $\hat{r}_0$ and $\hat{r}_1$ are estimators of the corresponding covariances computed from a discrete dataset). In all examples presented here, rate of convergence of the correlation function r_1/r_0 under the indirect observability determines the speed of convergence for both, the drift and the diffusion estimators. This can be shown analytically for the triad example and the model with the fast oscillating potential, but we rely on numerical evidence for the Truncated Burgers-Hopf model.

Under sub-sampling, the observational step is allowed to depend on the multiscale parameter, i.e. $\Delta \equiv \Delta(\epsilon)$. Nevertheless, the speed of convergence (with respect to ϵ) for the correlation function determines the leading order beahvior of the bias with respect to ϵ. The functional form of the estimators with respect to Δ determines how the sub-sampling step Δ enters the leading order term in the bias and thus, the interplay of these two issues is reflected in the conditions on the optimal sub-sampling strategy.

The behavior of the variance of approximating processes is quite different in the three examples presented in this paper. For the triad example, the estimator for the variance is not affected by indirect observability since the variance of the slow variable in the triad is identical to the variance of the reduced equation. For the example with the multiscale oscillating potential the variance of the approximating process depends on the multiscale parameter ϵ and it is possible to prove the convergence to the variance of the reduced equation as $\epsilon \to 0$ and estimate the speed of convergence with respect to ϵ. The Truncated Burgers-Hopf (TBH) model represent a more realistic example of a deterministic multiscale system with many degrees of freedom. Since the energy of this model is conserved, the stationary distribution has bounded support. The behavior of the variance can be estimated as the total number of variables is increased, but since the system is deterministic, all rigorous statements rely on the assumption of ergodicity and mixing. Nevertheless, the numerical data indicates that the TBH example falls in the category of equations suitable for homogenization. Therefore, our numerical investigation of the Truncated Burgers-Hopf example suggests that the sub-sampling strategy and the regression approach we have developed in a recent series of paper and illustrated here are applicable to a rather wide class of examples, including deterministic models.

Unfortunately, it is impossible to compute the L^2 speed of convergence for trajectories under the joint measure under the homogenization, since the process Y_t^ϵ is not explicitly obtained from X_t. Therefore, we concentrate on the behavior of the covariances, since they contain the information which is crucial for understanding the behavior of the estimators. In particular, we use the L^∞ speed of convergence of covariances as an indicator for speed of convergence of estimators, since the L^∞ norm is an upper bound for the L^2 norm. We demonstrate numerically that the rate of convergence as $\epsilon \to 0$ between correlation functions r_1/r_0 for the full and reduced models is consistent with the leading order term in the expression for the bias and, thus, the choice of sub-sampling step $\Delta \equiv \Delta(\epsilon)$ in the optimal sub-sampling schemes for the consistency of estimators.

In [5] we considered the vanishing lags estimators and emphasized that the behavior of the stationary lagged covariance functions for small lags determines the consistency criteria for the maximum likelihood estimators. Here we explore this approach further and we also show numerically that the leading order behavior of the bias can be inferred from the rate of convergence between correlation functions in the full

reduced equations. We demonstrate numerically that the leading order terms in the asymptotic expansion of the lagged covariance provide sufficient information to (i) understand the asymptotic behavior of the bias, (ii) construct the optimal sub-sampling regime to ensure the asymptotic consistency of the estimators, and (iii) construct the bias-corrected estimators for estimating parameters from a single dataset generated with a fixed, but unknown value of ϵ.

In contrast with the previous work, the goal of the present paper is to analyze the behavior of the Maximum Likelihood Parameter Estimators when the observable process, Y_t^ϵ, is generated by one of the multiscale systems discussed above and the process X_t is the limiting process derived as $\epsilon \to 0$ using the homogenization approach. For multiscale diffusions Y_t^ϵ, the homogenization implies the convergence of infinitesimal generators and is weaker than the L^p convergence of process trajectories. Hence, it is often a hard complementary step to establish rigorous bounds on the lagged covariance of the approximating process. Therefore, the main goals of this paper are the following
(i) analyze numerically the sub-sampling rates which enforce consistent parameters estimation, (ii) present efficient empirical arguments to deduce the optimal sub-sampling rates from the behavior of the correlation function near lag zero, and (iii) develop techniques for consistent parametric estimation using the data generated with a fixed, but unknown, multiscale parameter ϵ. We would like to emphasize that we only consider the vanishing lags regime when estimators are computed for small sub-sampling time-steps.

2.1. Estimation of the Ornstein-Uhlenbeck Parameters under Indirect Observability   In this paper we consider the problem of modeling a discrete dataset by a stochastic differential equation. Most parametric estimation techniques assume the direct observability, e.g. proposing a suitable parametric model based on the crude statistical properties of the data and directly estimating the unknown parameters by, for instance, constructing an approximate Likelihood function for the modeling SDE. In addition, various non-parametric techniques for estimating the drift and diffusion terms from discrete data (see, for example, [35, 8]) can be used in low-dimensional systems. Consequences of the multiscale nature of the data for non-parametric estimation are also considered in [9]. Due to the data limitations non-parametric techniques are applicable only if the dimension of the SDE is less than four. Here we consider the parametric approach under *indirect observability*, i.e. when the data generation mechanism involves multiscale effects which can potentially significantly affect the estimation procedure.

In particular, we study parametric estimation of the Ornstein-Uhlenbeck SDE

$$dX_t = -\gamma X_t dt + \sigma dW \qquad (2.3)$$

under *indirect observability*, i.e. when formal parameter estimators are constructed from the equation (2.3) as explicit functions $F(\Delta, N)$ of a discrete sub-sample $\{U_n = X_{n\Delta},\ 1 \le n \le N\}$, but the process X_t is not directly observable and the available data do not correspond to a discretely sampled Ornstein-Uhlenbeck trajectory. Instead, the data are sub-sampled at N discrete instants $t_n = n\Delta$ from an approximating process Y_t^ϵ such that $Y_t^\epsilon \to X_t$ as $\epsilon \to 0$. Particular examples of the process Y_t^ϵ will be presented in subsequent sections.

Parametric estimation of the OU SDE (2.3) from a discrete stationary dataset $U_n = X_{n\Delta}$, $n = 1 \dots N$ can be recast through differentiable functions of lagged covari-

ances $\{r_k(\Delta)\}_{k \in \mathbb{Z}}$, given by

$$r_k = r_k(\Delta) = \mathbb{E}[U_n U_{n+k}] = \frac{\sigma^2}{2\gamma} e^{-\gamma |k| \Delta}, \tag{2.4}$$

where $\mathbb{E}[\cdot]$ denotes the expected value with respect to the invariant distribution. In particular, parameters γ and σ in the OU SDE (2.3) can be expressed through the lagged covariances using the relationships $r_0 = (\sigma^2/2\gamma)$, $r_1 = r_0 \exp(-\gamma \Delta)$, which are readily inverted to yield the expressions

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

We define the estimators $\hat{\gamma}$, $\hat{\sigma}^2$ by

$$\hat{\gamma} = \hat{\gamma}(N, \Delta) = g(\hat{r}_0, \hat{r}_1), \quad \hat{\sigma}^2 = \hat{\sigma}^2(N, \Delta) = s(\hat{r}_0, \hat{r}_1), \tag{2.5}$$

where the covariances estimators are the *standard empirical covariance estimators* $\hat{r}_k$ given by

$$\hat{r}_k = \hat{r}_k(N, \Delta) = \frac{1}{N} \sum_{n=0}^{N-1} U_n U_{n+k}, \quad \text{for } k = 0, 1. \tag{2.6}$$

Estimators in (2.5) are asymptotically equivalent to the Maximum Likelihood Estimators for the OU SDE.

An important practical issue is to determine necessary and sufficient conditions for the consistency of parametric estimation under indirect observability, i.e. conditions to ensure that estimators $(\hat{\gamma}, \hat{\sigma}^2) \to (\gamma, \sigma^2)$ as $\epsilon \to 0$ when the unobservable random variables $U_n = X_{n\Delta}$ are replaced in (2.6), (2.5) by the observable variables $Y_{n\Delta}^\epsilon$. Typically, such conditions are formulated in terms of the number of observational points, $N \equiv N(\epsilon)$, the sub-sampling time-step, $\Delta \equiv \Delta(\epsilon)$, and the convergence parameter, ϵ.

In [5] we investigated the question of consistent estimation when the approximating process Y_t^ϵ is the Smoothed Ornstein-Uhlenbeck process

$$Y_t^\epsilon = \frac{1}{\epsilon} \int\limits_{t-\epsilon}^{t} X_s ds. \tag{2.7}$$

We obtained necessary and sufficient conditions for the consistency of the Maximum Likelihood Parameter Estimators based on adequately sub-sampled approximate data, and we investigated the optimal speed of convergence. In particular, the consistency conditions for vanishing lags estimators can be formulated as

$$\text{SOU Consistency}: \quad N(\epsilon) \to \infty, \quad \Delta(\epsilon) \to 0, \quad N(\epsilon)\Delta(\epsilon) \to \infty, \quad \epsilon/\Delta(\epsilon) \to 0. \tag{2.8}$$

If, for simplicity, we consider power-law relationships between N, Δ, and ϵ, then, the consistency requirement for the sub-sampling procedure becomes

$$\text{SOU Conditions}: \quad \Delta = \epsilon^\alpha, \quad \alpha \in (0,1), \quad N = \epsilon^{-\beta}, \quad \beta > \alpha. \tag{2.9}$$

We also derived explicit equation for the bias of the estimators and showed that for small sub-sampling step Δ, for small ϵ, and finite N the bias is approximately

$$\text{SOU Bias}: \quad \hat{\gamma} - \gamma \sim -\frac{\gamma \epsilon}{3\Delta} + \frac{C}{\sqrt{N\Delta}}, \quad \text{for } \Delta \geq \epsilon, \tag{2.10}$$

where C is some fixed constant. The second term in the right-hand side of (2.10) is a familiar term which describes errors due to a finite global observational time $T = N\Delta$. The first term on the right-hand side in (2.10) arises due to indirect observability of the data. Therefore, the last condition in (2.8) reflects the $O(\epsilon/\Delta)$ contribution to the bias under indirect observability and ensures that the bias converges to zero as $\epsilon \to 0$.

Often in practice ϵ is fixed and the number of observations can be chosen such that $\beta > 2 - \alpha$. The last condition is equivalent to the requirement $N\epsilon^2/\Delta \gg 1$. In this case, the second term $C/\sqrt{N\Delta}$ in (2.10) becomes negligible in comparison with the first term and the bias becomes

$$\text{SOU Bias:} \quad \hat{\gamma} - \gamma \sim -\frac{\gamma\epsilon}{3\Delta}, \quad \text{for } \Delta \geq \epsilon, \ \frac{N\epsilon^2}{\Delta} \gg 1. \tag{2.11}$$

As discussed previously, the scaling of the correlation function of Y_t^ϵ is directly related to the particular form of the consistency conditions in (2.8), (2.9) and the expression for the bias in (2.10), (2.11). Statistical properties of Y_t^ϵ can be computed explicitly and for small ϵ the correlation function near $\Delta = 0$ scales as

$$CF_{SOU}(\Delta) = \frac{\mathbb{E}\left[Y_t^\epsilon Y_{t+\Delta}^\epsilon\right]}{\mathbb{E}[(Y_t^\epsilon)^2]} \approx 1 - \frac{\Delta^2}{\gamma\epsilon} \quad \text{for} \ \ \Delta < \epsilon. \tag{2.12}$$

In particular, correlation scaling is directly linked to the condition $\alpha \in (0,1)$ in the consistency requirement (2.9) and the scaling of the bias in (2.11).

Convergence of the SOU process Y_t^ϵ to the OU process X_t does not fall in the homogenization category. Therefore, the goal of the present paper is to analyze the consistency conditions and the behavior of the bias when the approximating process is of the form (2.1) or (2.2). The L^p convergence of process trajectories cannot be easily established for such examples and we proceed by analogy by analyzing the scaling of the correlation functions for small lags. We show that the correlation function $CF(\Delta)$ scales differently compared with the $CF_{SOU}(\Delta)$ scaling for small lags and from the $CF(\Delta)$ scaling we deduce the appropriate condition for parameter estimation consistency and the scaling of the bias. We test our conjectured consistent sub-sampling schemes numerically for three multiscale systems which have been shown to yield a reduced effective stochastic differential equation under the homogenization procedure.

2.2. Modeling a given dataset generated at a fixed but unknown scale

In applied analysis of complex dynamic models, one is often confronted with the problem of estimating an effective stochastic model for the large-scale structures from a single dataset generated by intensive direct numerical simulations (DNS) of the full dynamics. Thus, the process Y_t^ϵ corresponds to the large-scale structures in a high-dimensional turbulent system (e.g. space-discretization of a turbulent PDE) and it is often known (or simply presumed) that the large scale dynamics Y_t^ϵ can be well approximated by a relatively low-dimensional parametrized stochastic system X_t. In this context the multiscale parameter ϵ often does not enter the equations explicitly and can be inferred indirectly by comparing the statistics of the large and small scales in full simulations. Moreover, the process Y_t^ϵ is non-Markov since it corresponds to a subset of all dynamic variables in the DNS.

Thus, in practice the value of ϵ is not known a priori, but, nevertheless, it is desirable to model the dataset with a specifically parametrized reduced stochastic model. Therefore, there is a discrepancy between the data and the model which can

potentially lead to significant errors in Maximum Likelihood Parameter Estimators. An important task is hence to develop a practical approach to compute accurate (unbiased) parameter estimators, when we are given a single dataset of indirect observations Y_t^ϵ corresponding to a fixed, but unknown value of the scale parameter ϵ. Here we outline a general approach to this problem; particular examples are presented in sections 3.2, 4.2, and 5.2.

We assume that the data are generated by a stochastic process Y_t^ϵ which is close to the unobserved process X_t in (3.3) in a suitable probabilistic metric, implying at least that statistical moments of the Y_t^ϵ are close to those of X_t. When X_t is the OU diffusion process, the drift and diffusion parameters can be expressed by an invertible function of the variance and the lagged correlation function of X_t. In [5] we showed that, for small lags, the correlation functions of the approximating process Y_t^ϵ and of the unobserved process X_t can be quite different. This can lead to significant errors when one implements naive parametric estimation for the parameters of the underlying SDEs.

When modeling a single dataset Y_t^ϵ generated as above for a fixed but unknown scale parameter ϵ, one has to determine from the data alone the correct sub-sampling regime corresponding to consistent estimation of the underlying diffusion parameters. To achieve this goal we let the sub-sampling time-step Δ vary, and consider multiple parameter estimators computed for several distinct values of Δ. Essentially, we consider estimators of the unknown parameters γ and σ^2 as functions $\hat{\gamma} \equiv \hat{\gamma}(\Delta)$ and $\hat{\sigma}^2 \equiv \hat{\sigma}^2(\Delta)$ of the observational time-step Δ. We show that the analysis of the graph of $\hat{\gamma}(\Delta)\Delta$ versus Δ clearly identifies the correct sub-sampling regime. Moreover, bias-corrected estimators can then be easily computed by linear regression of $\hat{\gamma}(\Delta)\Delta$ versus Δ. We expect that this methodology can be adequately extended to model discrete datasets generated by a reasonably wide range of complex dynamic systems. In particular, we expect that in many situations the behavior of $\hat{\gamma}(\Delta)$ and/or $\hat{\gamma}(\Delta)\Delta$ versus Δ will provide adequate information to recover the unbiased estimator of γ. This will be investigated in a subsequent paper.

The key observation for understanding the behavior of parameter estimators in this context is the behavior of the bias for small ϵ and Δ. In particular, we will show that for the parametric estimation of the Ornstein-Uhlenbeck under the indirect observability the leading term of the bias follows a hyperbolic relationship

$$\hat{\gamma}(\Delta) \approx \gamma + \frac{C(\epsilon)}{\Delta} \tag{2.13}$$

in the correct sub-sampling regime where the estimators are consistent under $\epsilon \to 0$ and a sufficient number of observations are used to compute the estimators, so that the term $1/\sqrt{N\Delta}$ is negligible compared to $C(\epsilon)/\Delta$. In (2.13), the term $C(\epsilon)$ does not depend on Δ, but only on ϵ. Note that this relationship is only valid in the correct sub-sampling regime.

In practice, the numerical data are generated by numerical simulations with a fine time-step δ. To study the estimators' behavior as we sub-sample the data, we consider several discrete sub-sampling time-steps $\Delta_j = j\delta$ and analyze the estimators $\hat{\gamma}(\Delta_j)$ and $\hat{\sigma}^2(\Delta_j)$ as functions of Δ_j. In sections 3.2, 4.2, and 5.2 we omit the index k in the notation for the sub-sampling time-step, but values of Δ in these sections are discrete and are integer multiples of the fine integration time-step δ.

3. Triad Model   The first model which we consider is the additive triad model
[27]

$$dx_t = \frac{1}{\epsilon}A_1 y_t z_t dt,$$

$$dy_t = \frac{1}{\epsilon}A_2 x_t z_t dt - \frac{1}{\epsilon^2}g_2 y_t dt + \frac{1}{\epsilon}s_2 dW_1(t), \qquad (3.1)$$

$$dz_t = \frac{1}{\epsilon}A_3 x_t y_t dt - \frac{1}{\epsilon^2}g_3 z_t dt + \frac{1}{\epsilon}s_3 dW_2(t),$$

where $A_1 + A_2 + A_3 = 0$, g_i, s_i are known parameters with g_2, $g_3 > 0$, W_1, W_2 are
independent Brownian motions, and $\epsilon > 0$ is the scale separation parameter. Here
the slow variable x_t plays the role of the approximating process introduced in the
generic presentation in section 2, i.e. $Y_t^\epsilon \equiv x_t$. The stationary covariance of x_t can be
computed explicitly for small lags Δ and is given by

$$CF_{x_t}(\Delta) = \frac{\mathbb{E}[x_t x_{t+\Delta}]}{\mathbb{E}[x_t^2]} \approx 1 - C\frac{\Delta^2}{\epsilon^2}, \qquad (3.2)$$

where $C = \gamma(g_2 + g_3)/2$ with γ given by (3.4). Details of the derivation are provided
in the appendix A.

   Asymptotic behavior of this model was analyzed in [27]. In particular, it was
shown that x_t can be treated as a slow variable compared to the fast y_t and z_t, and x_t
converges weakly to the OU process X_t as $\epsilon \to 0$. The limiting equations are obtained
by homogenization and are given by

$$dX_t = -\gamma X_t dt + \sigma dW_t, \qquad (3.3)$$

where W_t is Brownian motion, and γ and σ can be computed explicitly as

$$\gamma = \frac{-A_1}{2(g_2 + g_3)}\left(\frac{A_2 s_3^2}{g_3} + \frac{A_3 s_2^2}{g_2}\right), \qquad \sigma^2 = \frac{(A_1 s_2 s_3)^2}{2 g_2 g_3 (g_2 + g_3)}. \qquad (3.4)$$

As $\epsilon \to 0$, the weak convergence of x_t in (3.1) to the OU process can be proved using
the homogenization procedure [27, 33]. The derivation uses the asymptotic expansion
of the backward Kolmogorov equation. Weak convergence is equivalent to conver-
gence of expectations of bounded continuous functions of the process. Nevertheless,
we conjecture that the moments up to order 4 of x_t converge to the moments of the
Ornstein-Uhlenbeck process X_t in (3.3) as $\epsilon \to 0$. Indeed we note that the station-
ary distribution of the triad model can be computed explicitly for any ϵ (see [27]).
It can be shown that the stationary distribution for the triad is a product measure
and the stationary marginal distribution of x_t is Gaussian exactly matching the sta-
tionary distribution (also Gaussian) of the limiting Ornstein-Uhlenbeck process X_t.
Therefore, one-point moments of x_t and X_t are identical for all ϵ. Secondly, numerical
results confirming the convergence of the correlation function and the kurtosis (fourth
moment quantifying departures from Gaussian distributions)

$$Kurt(\tau) = \frac{\mathbb{E}\left[x^2(t)x^2(t+\tau)\right]}{\left(\mathbb{E}[x^2(t)]\right)^2 + 2\left(\mathbb{E}[x(t)x(t+\tau)]\right)^2} \qquad (3.5)$$

as $\epsilon \to 0$ are displayed in Figure 3.1. Therefore, for small ϵ, the consistent parametric
estimation procedure based on the observed x_t process should yield values for γ and σ
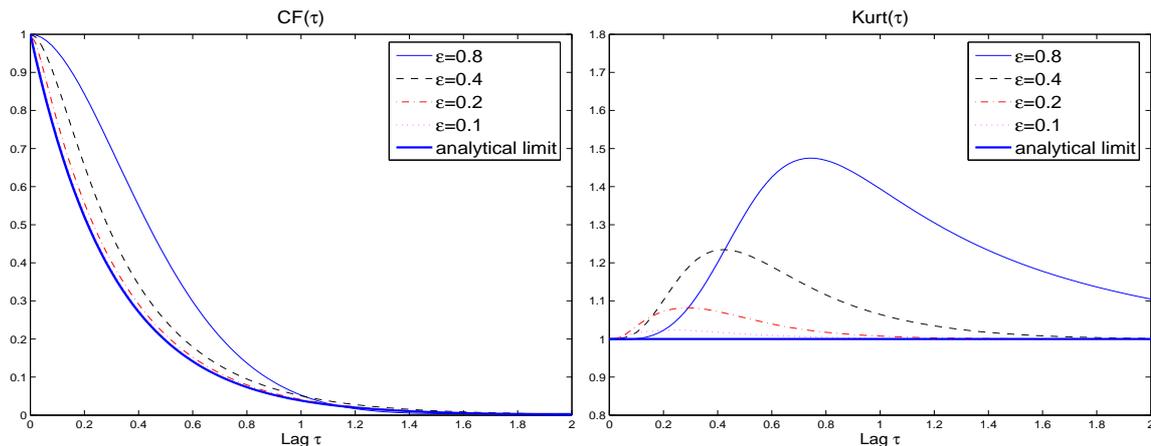
FIG. 3.1. *Left part: Convergence of the stationary correlation function $\mathbb{E}[x_t x_{t+\tau}]/\mathbb{E}[x_t^2]$ in the triad model (3.1) to the stationary correlation function $e^{-\gamma\tau}$ of the effective SDE (3.3) for several values of ϵ. Note: graphs for $\epsilon = 0.1$ and $\epsilon = 0.2$ overlap, thus graph for $\epsilon = 0.1$ is almost invisible on this figure. Right part: Convergence of the stationary kurtosis (3.5) in the triad model (3.1) to the stationary kurtosis of the effective SDE (3.3) $(Kurt(\tau) = 1)$ for several values of ϵ.*

close to the analytical asymptotic expressions in (3.4). Thus, we use expressions (3.4) to test the performance of the parameters estimation under indirect observability, when the observable data $x_t \equiv Y_t^\epsilon$ are generated by the triad model (3.1).

3.1. Sub-sampling for the Additive Triad Model

In this section we test various sub-sampling strategies for adaptive parametric estimation of the OU process X_t, when the estimators are computed from indirect observations, namely we use formulas (2.5), but covariances $\hat{r}_0$ and $\hat{r}_1$ are estimated from the data generated by the triad model, i.e. $U_n = x_{n\Delta(\epsilon)}$.

As $\epsilon \to 0$, we then compare the behavior of these estimators with the desired true parameter values, given by the explicit expressions (3.4). The triad model parameters in (3.1) are chosen to be

$$A_1 = 0.9, \ A_2 = -0.4, \ A_3 = -0.5,$$
$$g_2 = g_3 = 1,$$
$$s_2 = 3, \ s_3 = 5,$$

and we consider four values of ϵ

$$\epsilon = 0.1, \ 0.2, \ 0.3, \ 0.4.$$

With this choice of triad parameters values, the true values of the drift and diffusion coefficients of the unobservable limit process X_t are computed by formula (3.4), which yields the values

$$\gamma = 3.2625, \quad \sigma = 6.75. \tag{3.6}$$

The values of the "true" drift and diffusion parameters in (3.6) are only used to test the performance of the Maximum Likelihood Parameter Estimators under indirect

observability, i.e. when the data in (2.5) are generated by the slow variable x_t in the triad model with different sub-sampling strategies.

To determine the appropriate sub-sampling strategy for datasets generated by the triad equations we compare the asymptotic behavior of the correlation functions respective correlation functions CF_{SOU} and CF_{x_t} of the SOU and x_t processes (2.12) and (3.2), in order to adequately match the scaling of these two correlation functions with respect to the scale parameter ϵ. In particular, for small time lags Δ, the correlation function of the triad slow process x_t is given by (c.f. with (2.12))

$$CF_{\{x_t\}}(\Delta) \sim 1 - \frac{C\Delta^2}{\epsilon^2},$$

where C denotes a numerical constants. Since the scaling of the correlation function for the x_t in the triad model is different from the SOU process (ϵ^{-2} vs ϵ^{-1}), we conjecture that ϵ^2 (instead of just ϵ) should be substituted into the consistency conditions and the expression for the bias. Thus, when N, Δ are powers of ϵ the necessary condition for the consistency of the sub-sampling strategy for the triad model should be given by

$$\text{Triad Consistency:} \quad \Delta = \epsilon^{2\alpha}, \quad \alpha \in (0,1), \qquad N = \epsilon^{-2\beta}, \quad \alpha < \beta. \qquad (3.7)$$

Moreover, when N is large enough (we require a stronger condition $\beta > 2 - \alpha$ which implies that $N\epsilon^4/\Delta \gg 1$) the bias for the adaptive indirect parametric estimation from the triad data is proportional to $\gamma\epsilon^2/\Delta$, i.e.

$$\text{Triad Bias:} \quad \hat{\gamma} - \gamma \sim C\frac{\gamma\epsilon^2}{\Delta} \quad \text{for} \quad \frac{N\epsilon^4}{\Delta} \gg 1. \qquad (3.8)$$

In the simulations described below, the size of datasets is chosen to be $N = 500{,}000$ and is kept constant in all simulations; $N\epsilon^4/\Delta = 500 \gg 1$ holds for the smallest $\epsilon = 0.1$ and largest $\Delta = 0.1$ considered here. Therefore, the term on the right-hand side of (3.8) is the dominant term in the bias.

To support the above conjecture we compare several sub-sampling strategies. In particular, behaviors of the parameter estimators (2.5) computed from the triad data for four distinct sub-sampling strategies $\Delta = \epsilon$, $\Delta = \epsilon^2$, $\Delta = 4\epsilon^2$, and $\Delta = \epsilon^3$ are displayed in the left part of Figure 3.2. Sub-sampling strategies with $\Delta = \epsilon^{1.75}$, $\epsilon^{1.5}$, $\epsilon^{1.25}$ are also presented in the right part of Figure 3.2. As conjectured, as $\epsilon \to 0$, the bias is approximately constant for sub-sampling with $\Delta = \epsilon^2$ and $\Delta = 4\epsilon^2$, the bias decays to zero for sub-sampling with $\Delta = \epsilon^p$ with $p < 2$, and the bias grows unboundedly for sub-sampling with $\Delta = \epsilon^3$. Expression in (3.8) indicates that the bias should remain constant for both sub-sampling strategies $\Delta = \epsilon^2$ and $\Delta = 4\epsilon^2$, but errors for the sub-sampling strategy $\Delta = 4\epsilon^2$ should be four times smaller. Relative errors for $\hat{\gamma}$ for sub-sampling with $\Delta = \epsilon^2$ and $\Delta = 4\epsilon^2$ plotted in the lest part of Figure 3.2 support our conjecture in (3.8) to a very high precision. We also estimate the y-intercept for the straight line given by the relative errors computed from the sub-sampling strategy $\Delta = \epsilon$. The y-intercept for this line is approximately equal to 1.57, which means that this sub-sampling strategy should yield relative errors 1.57% as $\epsilon \to 0$. This is well within the numerical averaging errors for this particular example.

We also plot the error between the correlation functions $CF_{x_t}(\Delta)$ of the slow variables x_t in the triad model and the correlation function of the Ornstein-Uhlenbeck effective process $e^{-\gamma\Delta}$ with γ in (3.6) for several values of $\epsilon = 0.4, 0.3, 0.2, 0.1$ computed at a particular value of the lag $\Delta = 0.2$ in Figure 3.3. Similar results hold for other

values of the lag in the favorable sub-sampling regime $\Delta > \epsilon^2$. The error between the correlation functions clearly follows the ϵ^2 power law decay as $\epsilon \to 0$. This indicates that the lower bound for the speed of convergence $\hat{\gamma}(\epsilon) \to \gamma$ is equal to ϵ^2 for a fixed value of Δ. This supports our conjecture for the asymptotic behavior of the bias in (3.8). Therefore, numerical results for the convergence of correlation functions presented in Figure 3.3 clearly confirm our conjecture (3.7) about the optimal sub-sampling regime.

Moreover, it is possible to prove (see [27]) that the stationary distribution of the triad variables in (3.1) is a product Gaussian measure with mean zero and variance of $x_t \equiv Y_t^\epsilon$ exactly matching the variance of X_t in the reduced model (3.3) with (3.4), i.e. $Var\{x_t\} = Var\{X_t\}$ for all ϵ. Therefore, the estimator $\hat{r}_0$ in (2.5) (the variance estimate for the sub-sampled x_t dataset) does not depend on ϵ under indirect observability and the estimation error only depend on the length of observations. Thus, the leading order contribution to the bias arises due to the under-estimation of the lagged covariance $\hat{r}_1$ in (2.5). Therefore, this shows that sub-sampling errors manifest themselves first in $\hat{\gamma}$ and "propagate" into $\hat{\sigma}^2$ as evident from (2.5). Numerical convergence results for $\hat{\sigma}^2$ are indeed quite similar to those outlined for $\hat{\gamma}$, and hence are not presented here for brevity. We will sketch the behavior of $\hat{\sigma^2}$ for a different model.

3.2. Triad data analysis for fixed but unknown ϵ and bias-corrected estimators

In the previous section we used several datasets generated by numerical simulations of the multiscale triad system (3.1) with known values of ϵ to validate numerically the sub-sampling strategy (3.7) which guarantees that the bias converges to zero as $\epsilon \to 0$. In this section we use numerical data generated by the triad model (3.1) with a fixed particular value of ϵ to demonstrate how to model a specific dataset when the scale parameter ϵ is fixed but unknown.

When the data are generated by the SOU process, we have seen that the parameter estimators are linked to the true parameter values by an approximately hyperbolic relationship (2.13). By analogy with these results, we conjecture that similar approximately hyperbolic relationship holds in the correct sub-sampling regime for the data generated by the triad model. In particular, since the correct sub-sampling regime for the triad should be $\Delta > \epsilon^2$, we conjecture that the following approximate hyperbolic relationships should hold for the MLEs computed from the triad data

$$\hat{\gamma}(\Delta) \approx \gamma + \frac{C_1}{\Delta}, \qquad \hat{\sigma}^2(\Delta) \approx \sigma^2 + \frac{C_2}{\Delta} \quad \text{for} \ \ \Delta > \epsilon^2, \tag{3.9}$$

where C_1 and C_2 are unknown constants which do not depend on Δ and N is the number of observational points which is large enough (i.e. $N \gg \Delta/\epsilon^4$) so that the terms in (3.9) are the leading terms of the bias.

We use the triad data with $\epsilon = 0.3$ to elucidate the behavior of the estimators for a fixed small positive value of ϵ. In particular, the behavior of $\hat{\gamma}(\Delta)$ and $\hat{\gamma}(\Delta)\Delta$ as functions of Δ is presented in Figure 3.4. Numerical results displayed in the left part of Figure 3.4 show a typical hyperbolic graph for $\hat{\gamma}(\Delta)$, which is consistent with the expression (3.9). The critical scaling threshold for Δ is in this case $\epsilon^2 = 0.09$, but we point out that the estimator $\hat{\gamma}$ is biased for any value of Δ, including $\Delta > 0.09$. Moreover, plotting the function $\hat{\gamma}(\Delta)$ versus Δ for a given triad generated dataset for a fixed but unknown ϵ (left part of Figure 3.4) does not provide direct quantitative information about the unknown value of ϵ.

In order to deduce the value of the small scale parameter ϵ, we consider the graph of the function $\hat{\gamma}(\Delta)\Delta$ versus Δ. As conjectured, the graph of $\hat{\gamma}(\Delta)\Delta$ displayed in the right part of Figure 3.4 becomes approximately a straight line for $\Delta > 0.1$, with $\epsilon^2 = 0.09$ in this specific case. The region where $\hat{\gamma}(\Delta)\Delta$ does not follow the relationship (3.9) is clearly identifiable from Figure 3.4. Since the constants C_1 and C_2 in (3.9) are not known a-priori, it might be impossible to use this graphical information to compute the precise value of the multi-scale parameter ϵ. Nevertheless, the size of the region where the graph of $\hat{\gamma}(\Delta)\Delta$ does not follow an approximate straight line provides an estimate on the size of Δ for which the multiscale effects in datasets become significant.

Moreover, the expressions (3.9) suggest a practical approach for the construction of unbiased estimators. In particular, expressions (3.9) indicate that the graph of $\hat{\gamma}(\Delta)\Delta$ versus Δ should be an approximately straight line in the region $\Delta > \epsilon^2$ for the triad data. The slope of this straight line segment yields an unbiased estimator for the drift parameter γ. Linear regression is then an obvious standard tool to estimate the slope of the numerical graph of $\hat{\gamma}(\Delta)\Delta$ in the region $\Delta > \epsilon^2$.

To use the linear regression approach we eliminate the first few points in the graph of $\hat{\gamma}(\Delta)\Delta$, corresponding to small values of Δ; in our specific numerical example, we eliminate the two values $\Delta = 0.02, 0.06$ and compute the unbiased estimator for γ as the estimated slope coefficient for standard linear regression

$$\hat{\gamma}_{regression} \approx 3.23.$$

The relative error for the regression estimator is approximately 1%. This is a remarkable improvement compared to the standard MLE estimator $\hat{\gamma}$ under indirect observability. We point out that although the bias in $\hat{\gamma}$ decays as Δ^{-1}, the regression estimator is much more accurate than the estimator $\hat{\gamma}$ even for large values of Δ; for instance one has $\hat{\gamma}(\Delta = 0.58) \approx 3$.
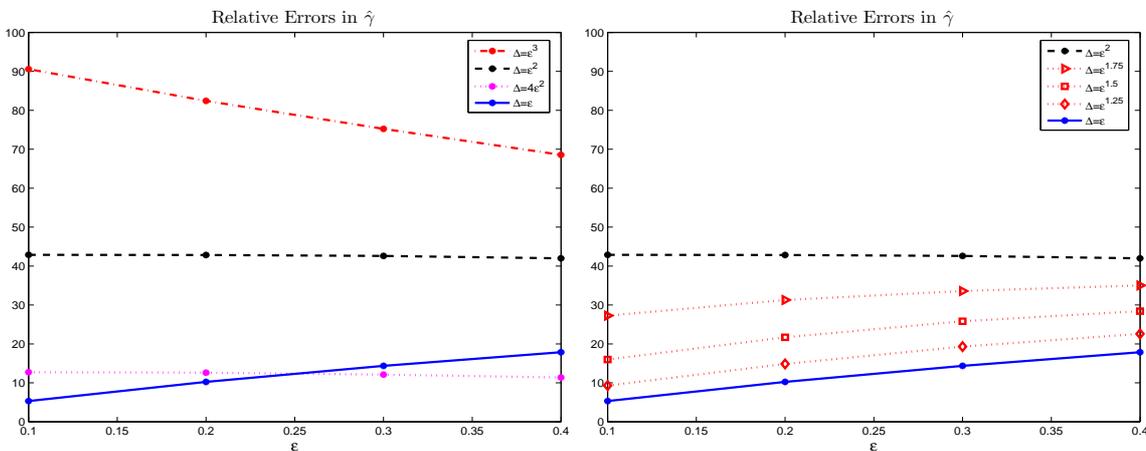
FIG. 3.2. *Relative errors (percent) for the estimator $\hat{\gamma}$ in* (2.5) *computed from the data of triad slow component x_t in the additive triad model in* (3.1) *sub-sampled with several different strategies. Left part: sub-sampling with $\Delta = \epsilon^3$, $\Delta = \epsilon^2$, $\Delta = 4\epsilon^2$, $\Delta = \epsilon$. Right part: sub-sampling with $\Delta = \epsilon^2$, $\Delta = \epsilon^{1.75}$, $\Delta = \epsilon^{1.5}$, $\Delta = \epsilon^{1.25}$, $\Delta = \epsilon$.*

Error in Correlation Function at lag=0.2
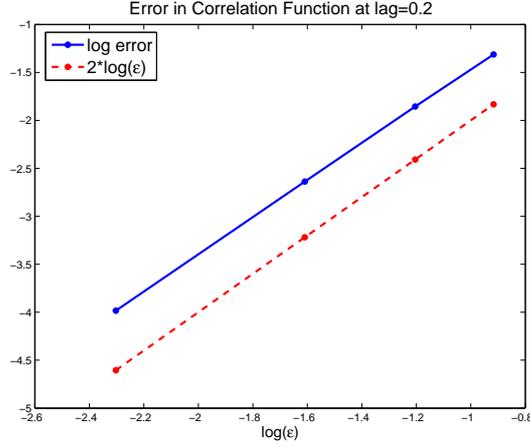
FIG. 3.3. *Log-log plot for the decay of the error between correlation function $|CF_{x_t}(\Delta) - CF_{X_t}(\Delta)|$ for $\epsilon = 0.4, 0.3, 0.2, 0.1$ computed at a particular lag $\Delta = 0.2$ where x_t is the slow variable in the triad model (3.1) and X_t is the Ornstein-Uhlenbeck process with parameters (3.6).*

$\hat{\gamma}$ $\hat{\gamma}(\Delta)\Delta$
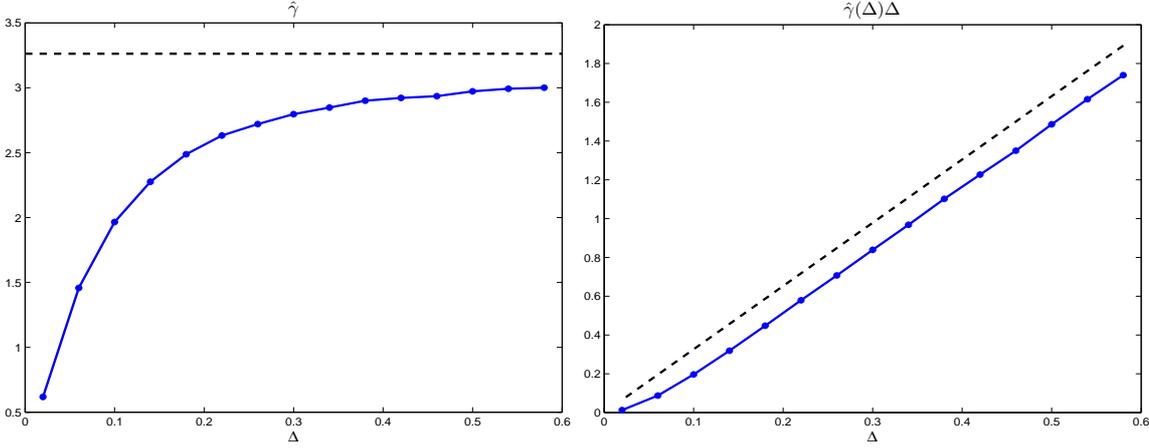
FIG. 3.4. *Left part: estimator $\hat{\gamma}(\Delta)$ for different values of Δ computed from the triad dataset with $\epsilon = 0.3$. Solid line: $\hat{\gamma}(\Delta)$ in (2.5), dashed line: analytical asymptotic value in (3.6) computed from (3.4). Right part: behavior of $\hat{\gamma}(\Delta)\Delta$ with $\hat{\gamma}$ computed for different values of Δ from the triad dataset with $\epsilon = 0.3$. Solid line: $\hat{\gamma}(\Delta)\Delta$, Dashed line: straight line with the slope $\gamma = 3.2625$ given by the analytical formula in (3.4).*

4. Multiscale Model with Fast Oscillating Potential

In this section we consider the parametric estimation of the Ornstein-Uhlenbeck process when the data are generated from a multiscale model of type (2.2). In particular, we consider the model discussed previously in [32]

$$dx_t = -gV'(x_t)dt - \frac{1}{\epsilon}p'\left(\frac{x_t}{\epsilon}\right)dt + \sqrt{2s}\,dB_t, \qquad (4.1)$$

where B_t is the Brownian motion, the large-scale potential is $V(x) = \frac{1}{2}x^2$, and the fast-oscillating part of the potential is $p(x) = \cos(x)$. The resulting SDE becomes

$$dx_t = -gx_t dt + \frac{1}{\epsilon}\sin\left(\frac{x_t}{\epsilon}\right)dt + \sqrt{2s}dB_t. \tag{4.2}$$

It has been shown (see [33] on homogenization for parabolic equations and references therein) that as $\epsilon \to 0$ the process x_t in (4.2) converges to the Ornstein-Uhlenbeck process

$$dX_t = -\gamma X_t dt + \sigma\, dW_t$$

with parameters γ and σ given by

$$\gamma = \alpha \frac{L^2}{Z\tilde{Z}}, \quad \sigma = \sqrt{2s\frac{L^2}{Z\tilde{Z}}}, \tag{4.3}$$

where L is the period of function $p(x)$ and hence equal to 2π and

$$Z = \int_0^L e^{\cos(y)/s}\,dy, \quad \tilde{Z} = \int_0^L e^{-\cos(y)/s}\,dy.$$

The convergence $x_t \to X_t$ is a weak convergence of generators which can be proven using homogenization techniques for parabolic equations. We would like to note that in this particular case $Z = \tilde{Z}$ due to the symmetry of the fast-oscillating potential.

The invariant density of the model (4.2) can be computed explicitly and is equal to

$$\rho_{inv}(x) = Ce^{-v(x)}, \quad \text{where} \quad v(x) = \frac{g}{2s}x^2 + \frac{1}{s}\cos\left(\frac{x}{\epsilon}\right), \tag{4.4}$$

where C is the normalization constant. Therefore, the invariant density has a fast-oscillating component, but converges to the Gaussian density weakly in the sense of test-functions.

We choose the parameter values in (4.2) as

$$g = 0.5, \quad s = 1, \tag{4.5}$$

and the corresponding homogenized coefficients are

$$\gamma = 0.3119, \quad \sigma = 1.117. \tag{4.6}$$

4.1. Sub-sampling Strategy for the Model with the Fast Oscillating Potential

In this section we test various sub-sampling strategies by computing estimates (2.5) for the Ornstein-Uhlenbeck parameters γ and σ from the data generated by the model with the fast oscillating potential in (4.2). Parametric estimation for the Ornstein-Uhlenbeck process and the sub-sampling issue when the data is generated by (4.2) was also investigated analytically and numerically in [32, 31]. Here we perform a more detailed numerical study of parametric estimation under indirect observability when the data is generated by the model (4.2). Numerical experiments support presented in this section support our conclusion about the crital sub-sampling scaling for parametric estimation under indirect observability when the limiting process is

obtain by a homogenization procedure. Moreover, our numerical experiments suggest that the conditions on the sub-sampling time-step Δ which guarantee consistent estimation can be extended compared with the analytical prediction in [31]; we discuss this issue in more detail in the conclusions.

First, to understand the behavior of the leading term in bias with respet to ϵ we analyze the rate of convergence of the correlation function $\hat{r}_1/\hat{r}_0$ computed from the data of x_t in (4.2). In particular, similar to the triad case we depict in Figure 4.1 the difference between the correlation function computed from for the process x_t in (4.2) and the correlation function for the Ornstein-Uhlenbeck process with parameters (4.6). Figure 4.1 clearly indicates that the correlation function for the process x_t in (4.2) converges with the same speed as the correlation function of the slow variable in the triad model (c.f. with the Figure 3.3). Since the stationary distribution (4.4) of the process x_t in (4.2) depends on the multiscale parameter ϵ, the variance of the process x_t in (4.2) depends on ϵ as well. It is possible (see B for an outline of the proof) to show explicitly the convergence of the variance, i.e. $Var\{x_t\} \to Var\{X_t\}$ as $\epsilon \to 0$. Moreover, it is also possible to analyze the rate of convergence with respect to ϵ in order to understand the behavior of the estimator $\hat{r}_0$ computed from the data x_t in (4.2). Using the proof sketched in the appendinx B one can derive that an upper bound for the speed of convergence is $|Var\{x_t\} - Var\{X_t\}| \le C\epsilon^2$. In fact, our numerical simulations indicate that the speed of convergence for the variance estimator is even faster. Thus, similar to the triad case, the convergence of the estimators $\hat{\gamma}$ and $\hat{\sigma}$ in (2.5) computed from the data of x_t in (4.2) is determined by the behavior of the correlation function $\hat{r}_1/\hat{r}_0$. Therefore, we conjecture that the critical scaling when the bias is constant with respect to ϵ for the estimators (2.5) computed by sub-sampling from the model with the fast oscillating potential (4.2) is identical to the triad case and is $\Delta = \epsilon^2$.
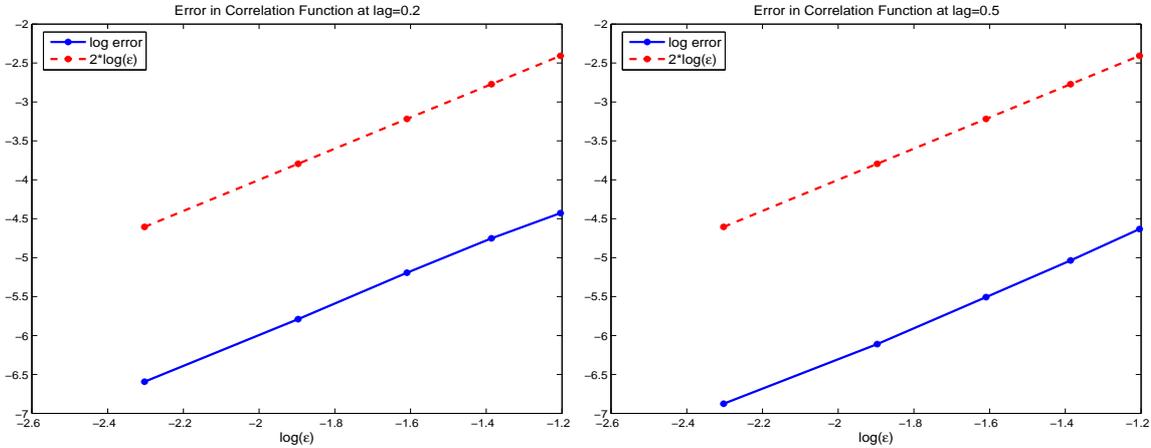
FIG. 4.1. *Log-log plot for the decay of the error between correlation function $|CF_{x_t}(\Delta) - CF_{X_t}(\Delta)|$ for $\epsilon = 0.3, 0.25, 0.2, 0.15, 0.1$ computed at a particular lag $\Delta = 0.2$ (left part) and $\Delta = 0.5$ (right part) where x_t is the process (4.2) and X_t is the Ornstein-Uhlenbeck process with parameters (4.6).*

Behavior of the bias in the estimator $\hat{\gamma}$ in (2.5) computed from the data of x_t in

(4.2) for several different sub-sampling strategies in presented in Figure 4.2. These results demonstrate that, indeed, the behavior of the estimators (2.5) computed using the data $U_n = x_{n\Delta}$, where x_t is a trajectory generated by the model with the fast oscillating potential (4.2) is identical to the behavior of the estimators computed from the data of the triad. In particular, the critical scaling $\Delta = \epsilon^2$ corresponds to the estimation when the error $|\hat{\gamma}(\Delta = \epsilon^2) - \gamma|$ remains constant with respect to ϵ. Similar to the triad case, estimation errors decay to zero for sub-sampling strategies $\Delta = \epsilon^p$ with $p < 2$. In particular, the relative errors for the scaling $\Delta = \epsilon$ follow a linear relationship and the intercept of this line computed by the linear regression is approximately 0.57 which corresponds to the estimated 0.57% relative error as $\epsilon \to 0$.

The total number of sample points N is much larger than Δ/ϵ^4 for all simulations depicted in Figure 4.2. In particular, the number of sample points for the largest $\Delta = 0.3$ is $N = 2 \times 10^6$, thus, $N\epsilon^4/\Delta \approx 600 \gg 1$ for the smallest value of $\epsilon = 0.1$ considered in the simultions. Therefore, calculations presented in this section confirm that the behavior of the bias for the estimators (2.5) computed from the data of x_t in the model with the fast oscillating potential (4.2) is

$$\hat{\gamma}(\Delta) - \gamma \sim C\frac{\epsilon^2}{\Delta}, \quad \text{for} \quad \frac{N\epsilon^4}{\Delta} \gg 1, \tag{4.7}$$

so that the number of observational points N in the dataset is large enough and the term $1/\sqrt{N\Delta}$ does not make a significant contribution.
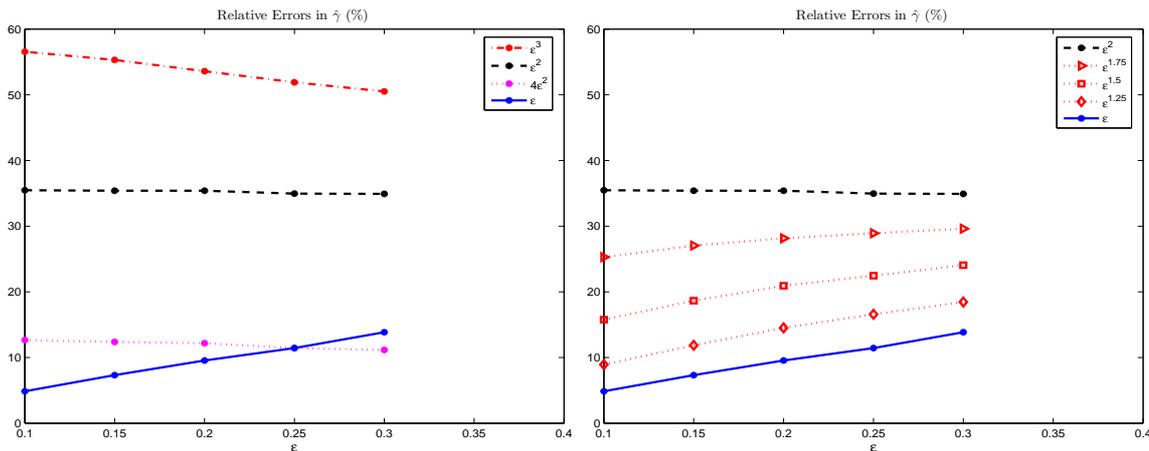
FIG. 4.2. *Relative errors (percent) in the estimator $\hat{\gamma}$ in (2.5) computed from the data of x_t generated by the model with the fast oscillating potential (4.2) sub-sampled with several different strategies. Left part: sub-sampling with $\Delta = \epsilon^3$, $\Delta = \epsilon^2$, $\Delta = 4\epsilon^2$, $\Delta = \epsilon$. Right part: sub-sampling with $\Delta = \epsilon^2$, $\Delta = \epsilon^{1.75}$, $\Delta = \epsilon^{1.5}$, $\Delta = \epsilon^{1.25}$, $\Delta = \epsilon$.*

4.2. Analysis of the Data generated by the Model with the Fast Oscillating Potential for fixed but unknown ϵ

We use the data generated by the model with the fast oscillating potential (4.2) with $\epsilon = 0.15$ to investigate the behavior of the estimators as a function of Δ. We assume that the value of ϵ is fixed but unknown and apply our regression technique to recover the correct value of the parameters. The left part of Figure 4.3 illustrates the

behavior of the estimator $\hat{\gamma}$ as a function of Δ. The behavior of $\hat{\gamma}(\Delta)$ vs Δ exhibits a hyperbolic profile consistent with the estimate for the bias in (4.7). The overall behavior is similar to the triad case, except the damping parameter, $\hat{\gamma}$ is severely overestimated for small Δ. In particular, the "true" homogenized value of the damping parameter is $\gamma = 0.3119$ and values of the estimator $\hat{\gamma}(\Delta)$ for $\Delta \in [0.002, \ldots, 0.22]$ are in the range $[0.49, \ldots, 0.326]$. This is the consequence of the fact that the correlation function for x_t in the potential model decays faster than the homogenized correlation function $e^{-\gamma\tau}$ for small lags.

We plot the behavior of $\hat{\gamma}(\Delta)\Delta$ vs Δ in the right part of Figure 4.3. The graph $\hat{\gamma}(\Delta)\Delta$ vs Δ becomes approximately a straight line between $\Delta = 0.02$ and $\Delta = 0.04$ providing an approximate indication for the range of the multiscale effects in the data ($\epsilon^2 = 0.0225$ in this case). Therefore, we can also use the regression estimator to estimate the effective damping parameter from the slope of the line $\hat{\gamma}(\Delta)\Delta$ vs Δ where $\hat{\gamma}(\Delta)$ is computed from the dataset generated by the potential model (4.2) with $\epsilon = 0.15$. It is clearly visible from the graph of $\hat{\gamma}(\Delta)\Delta$ vs Δ that the first two points for $\Delta = 0.002$, 0.01 do not follow the linear relationship and should be neglected in the regression estimator. This also can be quantified numerically by computing the goodness of fit for the linear regression of the graph $\hat{\gamma}(\Delta)\Delta$ vs Δ on different intervals for the sub-sampling parameter Δ. Thus, the regression estimator discussed in section 2.2 computed using the data $\hat{\gamma}(\Delta)$ with $0.02 < \Delta \leq 0.22$ becomes

$$\hat{\gamma}_{regression}(\epsilon = 0.15) \approx 0.3139$$

which is in a very good agreement (only 0.6% relative error) with the analytical prediction for the homogenized coefficient (4.6). We would like to point out that if a straightford estimation is used, the estimator $\hat{\gamma}(\Delta = 0.22) = 0.326$ which amounts to the 4.5% relative error. Therefore, the regression estimator significantly outperforms the straightforward estimation in the indirect observability context. Moreover, the regression estimator computed on the interval $0.02 < \Delta \leq 0.22$ is much more accurate than the standard estimator $\hat{\gamma}(\Delta = 0.22)$, but the regression estimator is computed using the same number of observaional point, N, as the standard estimator $\hat{\gamma}(\Delta = 0.22)$.

Overall, parametric estimation of the Ornstein-Uhlenbeck process under indirect observability using the data generated by the SDE with the fast oscillating potential (4.2) falls in the same category as the triad model discussed in section 3. Our numerical simulations clearly indicate that the speed of convergence of the correlation function r_1/r_0 in the multiscale SDE (4.2) controls the speed of convergence (w.r. to ϵ) of the estimator for the drift parameter $\hat{\gamma}$ and, also, determines the leading term in the bias with respect to ϵ. Our numerical simulations also indicate that the behavior of the estimator for the diffusion $\hat{\sigma}^2$ is also controlled by the behavior of the correlation function and, although, the stationary variance of the process x_t in the SDE (4.2) depends on ϵ, it does not significantly contribute to the scaling behavior of the estimator for the diffusion parameter $\hat{\sigma}^2$. Moreover, the regression technique outlined in section 2.2 can be successfully applied to datasets generated by the SDE (4.2) with a fixed but unknown value of ϵ.

5. Truncated Burgers-Hopf Equations   In this section we consider the sub-sampling problem for the Truncated Burgers-Hopf (TBH) model

$$\partial_t U_\Lambda(x,t) + \frac{1}{2}\partial_x P_\Lambda U_\Lambda^2(x,t) = 0, \tag{5.1}$$
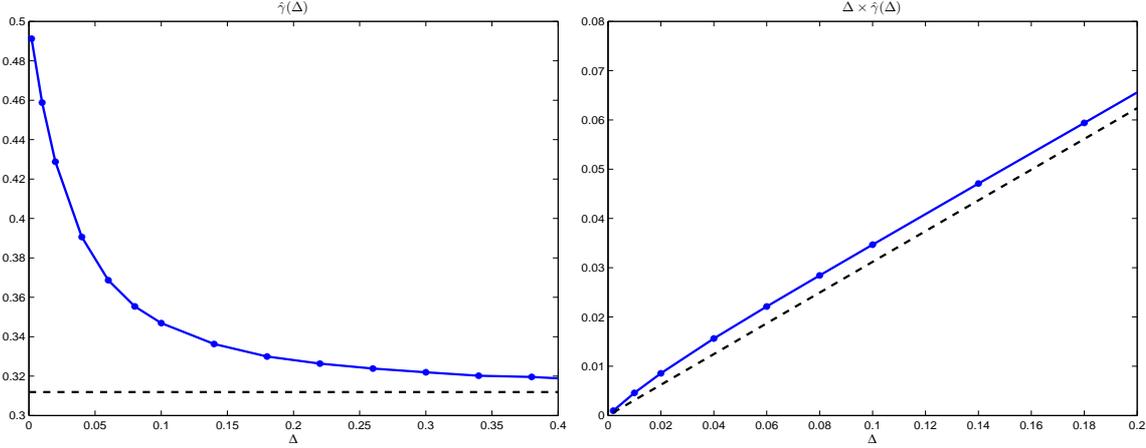
FIG. 4.3. *Left part: estimator $\hat{\gamma}(\Delta)$ for different values of Δ computed from the data generated by the SDE with the fast oscillating potential (4.2) with $\epsilon = 0.15$. Solid line: $\hat{\gamma}(\Delta)$ in (2.5), dashed line: analytical asymptotic value in (4.3) computed from (4.5). Right part: behavior of $\hat{\gamma}(\Delta)\Delta$ with $\hat{\gamma}$ computed for different values of Δ from the data generated by the SDE with the fast oscillating potential (4.2) with $\epsilon = 0.15$. Solid line: $\hat{\gamma}(\Delta)\Delta$, Dashed line - straight line with the slope $\gamma = 0.3119$ given by the analytical formula in (4.3).*

where $x \in [0, 2\pi]$ and P_Λ is the projection operator in Fourier space

$$P_\Lambda u(x,t) = \sum_{|k| \leq \Lambda} u_k(t) e^{ikx}, \tag{5.2}$$

and U_Λ is a finite dimensional projection

$$U_\Lambda(x,t) = \sum_{|k| \leq \Lambda} u_k(t) e^{ikx} = P_\Lambda u(x,t). \tag{5.3}$$

Equation (5.3) is supplemented with the reality condition $u_{-k}(t) = u_k^*(t)$, where u^* denotes complex conjugation. Equation in (5.1) can be recast as a 2Λ-dimensional system of ordinary differential equations

$$\frac{d}{dt} u_k = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|,|q| \leq \Lambda}} u_p^* u_q^*, \quad |k| \leq \Lambda, \tag{5.4}$$

where $u_k \equiv u_k(t)$. Equations in (5.4) conserve energy

$$E = \frac{1}{4\pi} \int U_\Lambda^2 dx = \sum_{k=1}^{\Lambda} |u_k|^2. \tag{5.5}$$

and Hamiltonian

$$H = \frac{1}{12\pi} \int_0^{2\pi} U_\Lambda^3 dx.$$

This model was introduced in [23, 24] and statistical properties of this model were further studied in [1]. Stochastic modeling for the large-scale structures in the Truncated Burgers-Hopf is a more realistic problem since the model arises as a suitable projection of a partial differential equation. Thus, we can expect that the effects observed in the stochastic models of the TBH dataset are typical for datasets generated by numerical integration of realistic partial differential equations.

Since the equation for u_0 is trivial, we can assume $u_0(0) = 0$ without loss of generality. In particular, it was demonstrated that generic initial conditions correspond to $H \approx 0$. Moreover, for generic initial conditions Fourier coefficients achieve equipartition and equilibrium statistical properties of Fourier coefficients follow a joint Gaussian distribution

$$\pi(u_1, u_2, \ldots u_\Lambda) = C e^{-\beta E}, \tag{5.6}$$

where C is a normalization constant, E is the energy defined in (5.5), and β is the inverse of the temperature determined by the energy of the initial condition. According to the distribution in (5.6) modes u_k achieve equipartition of energy with

$$Var\{\text{Re } u_k\} = Var\{\text{Im } u_k\} = \frac{1}{2\beta}.$$

In [29] it was demonstrated that homogenization can be applied to the TBH model. In particular, equations in (5.4) can be modified by introducing a small parameter, ϵ, into the model and the limit as $\epsilon \to 0$ of the modified equations can be computed explicitly. The approach to introduce the small scale parameter ϵ depends specifically on the number of essential variables in the equation, i.e. the number of modes to be retained in the limit. For a detailed discussion on homogenization of the TBH model see [29]. Here we only consider the case where u_1 is the essential variable. The rest of the variables $k = 2 \ldots \Lambda$ are considered to be fast and are eliminated by the homogenization procedure.

Effective Equations for u_1. When u_1 is the only essential variable, the TBH equations are modified as follows

$$\dot{u}_1 = -\frac{i}{2\epsilon} \sum_{\substack{p+q+1=0 \\ 2 \le |p|, |q| \le \Lambda}} u_p^* u_q^*, \tag{5.7}$$

$$\dot{u}_k = -\frac{ik}{2\epsilon}(u_{k+1}u_1^* + u_{k-1}u_1) - \frac{ik}{2\epsilon^2} \sum_{\substack{k+p+q=0 \\ 2 \le |p|, |q| \le \Lambda}} u_p^* u_q^*, \qquad (k \ge 2) ,$$

where we used the reality condition $u_{-k} = u^*$ to simplify the right-hand side of the second equation.

The limiting behavior of u_1 is given by the Ito SDE

$$da_k = B(a)a_k dt + H(a)a_k dt + \sqrt{2}\sigma(a)dW_k(t), \qquad (k = 1, 2) \tag{5.8}$$

where $a = (a_1, a_2) \equiv (\text{Re } u_1, \text{Im } u_1)$ and

$$
\begin{cases}
B(a) = -(1 - 2M^{-1})\mathcal{E}^{-1}(a)\left(\mathcal{E}^{1/2}(a)I_2|a|^2 + \mathcal{E}^{3/2}(a)I_f\right), \\[2mm]
H(a) = -M^{-1}\mathcal{E}^{-1/2}(a)|a|^2 I_2 + 2\mathcal{E}^{1/2}(a)I_2 - 3M^{-1}\mathcal{E}^{1/2}(a)I_f, \\[2mm]
\sigma^2(a) = \mathcal{E}^{1/2}(a)|a|^2 I_2 + \mathcal{E}^{3/2}(a)I_f, \\[2mm]
\mathcal{E}(a) = M^{-1}(E - |a|^2).
\end{cases}
\tag{5.9}
$$

Details of the derivation are provided in [29]. We point out that the proof of convergence requires assumptions of ergodicity and mixing on the deterministic system in (5.7). Expressions in (5.9) depend on the following 4 parameters: $M = 2\Lambda - 2$ is the number of fast degrees of freedom, E is the total energy of the full TBH model, I_2 and I_f are two parameters related to the averaged cross-correlations of the fast variables, i.e.

$$
\begin{aligned}
I_2 &= I[\text{Re } u_2, \text{Re } u_2] = I[\text{Im } u_2, \text{Im } u_2], \\
I_f &= I[f^r, f^r] = I\left[f^i, f^i\right],
\end{aligned}
\tag{5.10}
$$

where $I[\cdot, \cdot]$ is a short-hand notation for the area under the graph of a correlation function and f^r and f^i are the terms without u_2 on the right-hand-side of u_1^{re} and u_1^{im}, respectively. For specific details see [29].

Parameters I_2 and I_f are estimated from a single microcanonical simulation of the full equations on the energy surface $E = 2\Lambda$, so that $\langle (\text{Re } u_k)^2 \rangle_t = \langle (\text{Im } u_k)^2 \rangle = 1$. Then, I_2 and I_f depend only on the truncation size and can be computed a priori for all initial conditions for a given truncation. For the process $a_1 = \text{Re } u_1$ generated by the TBH dynamic model, numerical convergence of empirical estimates of the stationary correlations and kurtosis defined by (3.5) are presented in Figure 5.1.

Equations (5.8) are highly nonlinear, but can be simplified by considering the limit of infinitely many fast variables. In particular, in the limit $M \to \infty$ equations in (5.8) become

$$
da_k = b(a)a_k dt + \sqrt{2}s(a)dW_k, \qquad k = 1, 2,
\tag{5.11}
$$

where the drift and the diffusion simplify to

$$
\begin{aligned}
b(a) &= -\left(\sqrt{2\beta}|a|^2 I_2 + (2\beta)^{-1/2}I_f - \sqrt{2/\beta}I_2\right) \\
s^2(a) &= (2\beta)^{-1/2}|a|^2 I_2 + (2\beta)^{-3/2}I_f.
\end{aligned}
$$

We would like to point out that the limits $\epsilon \to 0$ and $M \to \infty$ do not commute, and (5.11) is only an approximation for the reduced model in (5.8). In particular, values of I_2 and I_f depend on M, but are treated as fixed in the limit of (5.8) as $M \to \infty$.

It is straightforward to verify that the Gaussian density

$$
\rho(a) = \frac{1}{2\pi\beta}e^{-\beta|a|^2}
\tag{5.12}
$$
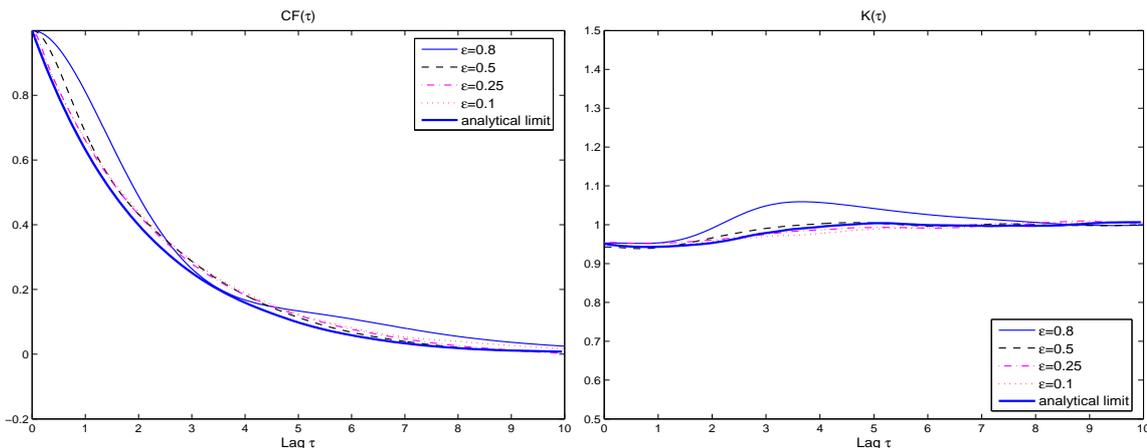
FIG. 5.1. *Left part: Convergence of the stationary correlation function $\langle u_1^{re}(t)u_1^{re}(t+\tau)\rangle$ in the modified TBH model (5.7) to the stationary correlation function of a_1 in the effective equation in (5.8) for several values of ϵ. Note: graphs for $\epsilon = 0.1$ and $\epsilon = 0.25$ nearly overlap, thus graph for $\epsilon = 0.1$ is almost invisible on this figure. Right part: Convergence of the stationary kurtosis (3.5) for $u_1^{re}(t)$ in the modified TBH model (5.7) to the stationary kurtosis of a_1 in the effective equation (5.8) for several values of ϵ.*

defines a stationary distribution for the equations in (5.11). Equation (5.11) have cubic and linear damping and multiplicative and additive noises. In [29] equations (5.7) and (5.8) were considered in the weak energy regime $\beta = 50$ ($Var\{a_1\} = Var\{a_2\} = 0.01$). Here we consider the truncation size $\Lambda = 20$, so that the total energy is $E = 0.4$ and the number of fast variables is $M = 2\Lambda - 2 = 38$. Parameters I_2 and I_f become

$$I_2 = 0.14, \ I_f = 4.3. \tag{5.13}$$

Although the reduced equations in (5.8) and (5.11) are valid for any energy level (arbitrary β), they are close to a linear OU system ($I_2 \ll I_f$) for $\beta = 50$. Moreover, numerical simulations indicate that higher moments are approximately Gaussian; this also indicates that for $\beta \gg 1$, the cubic terms and multiplicative noises become weak. Therefore, in the regime $I_2 \ll 1$, the modes $(a_1, a_2) = (\text{Re } u_1, \text{Im } u_1)$ become approximately uncorrelated and the two-point correlation function is well-approximated by an exponential function. Stationary correlation functions of a_1 in simulations of the effective SDE (5.8) and (5.11) are compared to the exponential function $e^{-(2\beta)^{-1/2}I_f t}$ in the left part of Figure 5.2. The kurtosis for the distribution of $a_1 = \text{Re } u_1$, defined as in (3.5), is estimated by simulations of the reduced models (5.8) and (5.11) is presented in the right part of Figure 5.2. Kurtosis (see definition (3.5)) quantifies non-Gaussianity since $Kurt(\tau) = 1$ for Gaussian processes. The observed kurtosis behavior indicates that both reduced models in (5.8) and (5.11) are practically Gaussian from the point of view of second, third, and fourth moments. This represents a realistic situation when the non-Gaussian features of the dataset are barely detectable and cannot be estimated reliably.

Therefore, in the weak energy regime $\beta \gg 1$, the system (5.11) is well-approximated by the reduced system with $I_2 = 0$. In this regime the reduced equation
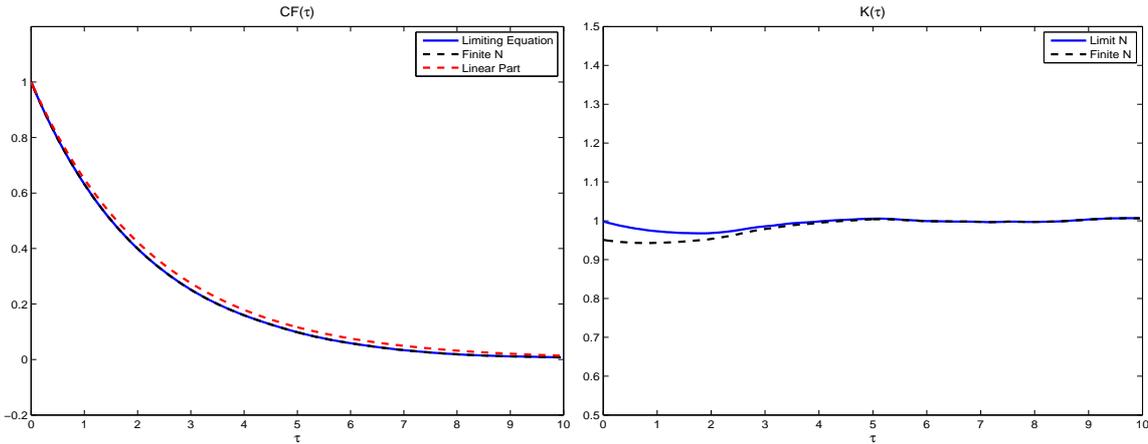
FIG. 5.2. *Left part: correlation function of $a_1 = \mathrm{Re}\, u_1$ in the simulations of the effective model in (5.8) (black dashed line), limiting equations in (5.11) (blue solid line), and the exponential function $e^{-(2\beta)^{-1/2}I_f t}$. Right part: Kurtosis $Kurt(\tau)$ in the simulations of the reduced models (5.8) and (5.11).*

becomes a system of two independent linear Ornstein-Uhlenbeck processes

$$da_k = -\gamma a_k + \sqrt{2}s dW_k, \quad k = 1,2 \tag{5.14}$$

with $\gamma = (2\beta)^{-1/2}I_f$ and $s^2 = (2\beta)^{-3/2}I_f$. It is easy to verify that the linear system (5.14) has a Gaussian stationary distribution given by (5.12).

In order to evaluate in this context the performance of the parameters estimators (2.5) we consider that as $\epsilon \to 0$, the limiting "true" value of the drift and diffusion coefficients in the OU approximation of the modified TBH model are

$$\gamma = (2\beta)^{-1/2}I_f, \quad \sigma^2 = 2s^2 = 2(2\beta)^{-3/2}I_f. \tag{5.15}$$

We fix the value of I_f as in (5.13) and $\beta = 50$. The coefficients γ and σ^2 then become

$$\gamma = 0.43, \quad \sigma^2 = 0.0086.$$

In particular, we expect that in the correct sub-sampling regime, and as $\epsilon \to 0$, the empirical estimators $\hat{r}_0$ and $\hat{r}_1$ should converge to the drift and diffusion parameter values given by (5.15). We point out that additional assumptions have been made in the derivation of the equation (5.8). The main simplifying assumption is that the cross-correlation is zero between $a_1 = \mathrm{Re}\, u_1$ and $a_2 = \mathrm{Im}\, u_1$. It can be shown numerically that this cross-correlation is very close to zero, but there is no analytical justification for this point. The assumption of zero cross-correlation leads to small discrepancies between the modified model (5.7) and the reduced model (5.8). Moreover, the weakly non-Gaussian nature of the full model can also lead to small discrepancies in the estimation procedure. Nevertheless, parametric estimation of the linear OU model from the TBH data represents a more realistic case when the proposed parametric model is not perfect even in the limit $\epsilon \to 0$. We demonstrate that our conclusions about sub-sampling rates and unbiased estimators are still valid in this case.

5.1. Sub-sampling Strategy for the Truncated Burgers-Hopf Model

We test the performance of the parameter estimators (2.5) for various sub-sampling strategies by studying four TBH datasets generated by simulating the dynamics (5.7) for four values of ϵ. Since the effective linear dynamic equations in (5.14) decouple a_1 and a_2, we only need to study the parameter estimators for the dynamics of a_1. Parametric estimators for the dynamics of a_2 have similar behavior. To test the sub-sampling strategy, we consider the dataset $U_n = \mathrm{Re}\, u_1(n\Delta)$ and compare the limit behavior of the parameter estimators (2.5) as $\epsilon \to 0$ with the analytical values given by (5.15).

Since a similar homogenization procedure is applied to derive the reduced equations in the Triad model and the modified TBH model, the time-scale of the fast variables in both models is $O(\epsilon^2)$. Hence the behavior of the estimators $\hat{\gamma}$ and $\hat{\sigma^2}$ should be similar to the Triad case. We study again several sub-sampling strategies depicted in Figure 5.3. Relative errors for the estimator $\hat{\gamma}$ exhibit the same behavior as for the Triad data (cf Figures 5.3 and 3.2). When Δ is a power of ϵ, the critical scaling is clearly $\Delta = \epsilon^2$ so that the estimation error in $\hat{\gamma}$ is approximately constant for different values of ϵ. Figure 5.3 shows that relative errors in the estimation of parameters decay to zero as $\epsilon \to 0$ only for sub-sampling regimes where $\Delta(\epsilon) = \epsilon^{2\alpha}$ with $\alpha \in (0, 1)$.

5.2. Unbiased parameter estimators for fixed but unknown ϵ

As in section 3.2, we also analyze how the estimator $\hat{\gamma}$ depends on Δ for an arbitrary fixed dataset generated by simulations of the modified TBH model with a particular fixed value of ϵ. In particular, for a fixed $\epsilon = 0.3$ we plot the graphs of $\hat{\gamma}(\Delta)$ and $\hat{\gamma}(\Delta)\Delta$ as functions of the sampling time-step Δ. These 2 graphs are presented in Figure 5.4, where the left part displays the graph of $\hat{\gamma}(\Delta)$ as a function of the sub-sampling time step, Δ. As for the Triad model, the graph of $\hat{\gamma}(\Delta)$ vs Δ provides only limited information about the estimator bias or the appropriate value of Δ for consistent parameter estimation. This graph shows nevertheless that the estimator $\hat{\gamma}(\Delta)$ is sensitive to changes in Δ even for very large $\Delta \approx 0.5$.

On the other hand, the graph of $\hat{\gamma}(\Delta)\Delta$ versus Δ (see the right part of Figure 5.4) provides much better estimates for the unknown value of the small scale parameter ϵ and for the favorable sub-sampling regime. The graph of $\hat{\gamma}(\Delta)\Delta$ versus Δ becomes approximately a straight line for $\Delta > 0.15$. Therefore, $\Delta \approx 0.15$ is correctly identified as the time-scale of fast variables in this problem. Small discrepancy between $\epsilon^2 = 0.09$ and 0.15 can be attributed to the finite-ϵ effects since the reduced model for the accelerated TBH system is slightly different from the Ornstein-Uhlenbeck process and should include small nonlinear terms.

To construct an unbiased estimator of γ, we use a linear regression fit to the graph of $\hat{\gamma}(\Delta)\Delta$, for $\Delta > 0.15$. The slope coefficient generated by this linear regression becomes a new estimate of γ, given by

$$\hat{\gamma}_{regression}(\epsilon = 0.3) \approx 0.45.$$

The relative error for this new regression estimator is approximately 5% since the "true" value is $\gamma \approx 0.43$. The regression estimator performs much better than the biased estimator $\hat{\gamma}$, since the relative errors for the biased estimator are bigger than 10% for all values of Δ.

We also illustrate the behavior of $\hat{\gamma}(\Delta)$ and $\hat{\sigma}^2(\Delta)$ as functions of Δ for another simulated TBH dataset computed for $\epsilon = 0.1$. Overall, the observed numerical behavior of the parameter estimators is similar to the behavior presented for $\epsilon = 0.3$.

In particular, $\hat{\gamma}(\Delta)\Delta$ and $\hat{\sigma}^2(\Delta)\Delta$ are plotted in Figure (5.5). There is a noticeable change in the curvature of both graphs at $\Delta \approx 0.01$. This value of Δ corresponds to the critical scaling $\Delta = \epsilon^2$. Therefore, the fast time-scale is clearly identifiable on both graphs. Moreover, linear regression provides the following values for the slopes of the two approximating line segments associated to these two graphs

$$\hat{\gamma}_{regression}(\epsilon = 0.1) \approx 0.4101, \quad \hat{\sigma}^2_{regression}(\epsilon = 0.1) \approx 0.0089. \tag{5.16}$$

Similar to regression estimators considered for other models in this paper, we neglect the first five points which do not follow the expected linear relationship. We plot the biased estimator (2.5) and the linear regression estimator computed from the TBH dataset with $\epsilon = 0.1$ on the same display in Figure (5.6). The relative errors for the regression estimators for γ and σ^2 are 5% and 3%, respectively. On the other hand, we would like to emphasize that the regression estimators are extremely robust and are not affected by the necessity to choose a particular fixed value of the sub-sampling time-step, Δ.
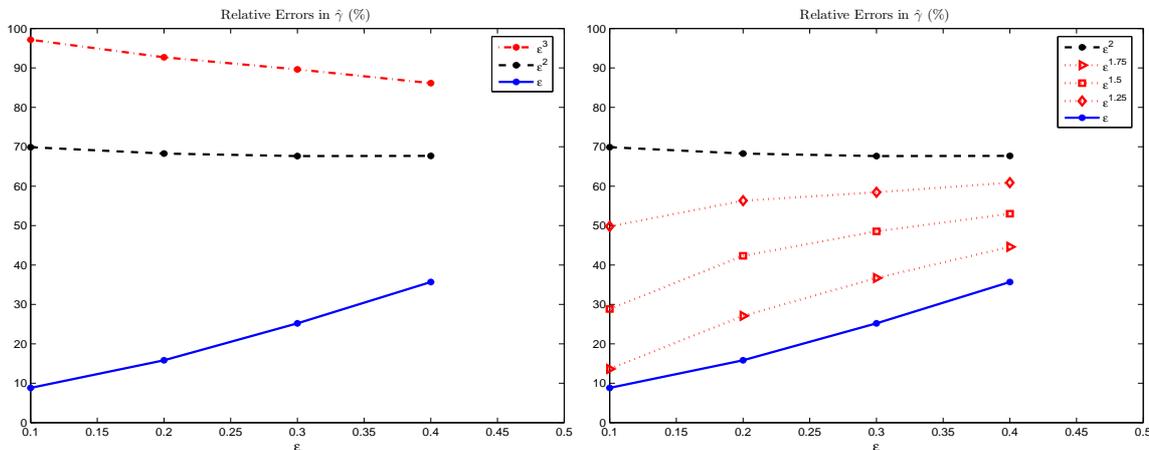
FIG. 5.3. *Relative errors (%) in the estimator $\hat{\gamma}$ in (2.5) computed from the data of x_t in the Truncated Burgers-Hopf model in (5.7) sub-sampled with several different strategies. Left part: sub-sampling with $\Delta = \epsilon^3$, $\Delta = \epsilon^2$, $\Delta = \epsilon$. Right part: sub-sampling with $\Delta = \epsilon^2$, $\Delta = \epsilon^{1.75}$, $\Delta = \epsilon^{1.5}$, $\Delta = \epsilon^{1.25}$, $\Delta = \epsilon$.*

6. Conclusions

In this paper we have studied numerically three examples of parametric estimation of an effective SDE for the process X_t from approximate data when the observations are sub-sampled from a discrete dataset generated by an approximating multiscale process Y_t^ϵ. We considered cases when the approximating process Y_t^ϵ converges to the process X_t as $\epsilon \to 0$ which allows to test our estimation strategy. The main aspect here is parametric estimation under *indirect observability*, i.e. when the available numerical data are not generated by the limit process X_t, but instead, are generated by an approximate dynamic process Y_t^ϵ. This is a typical situation in many physical
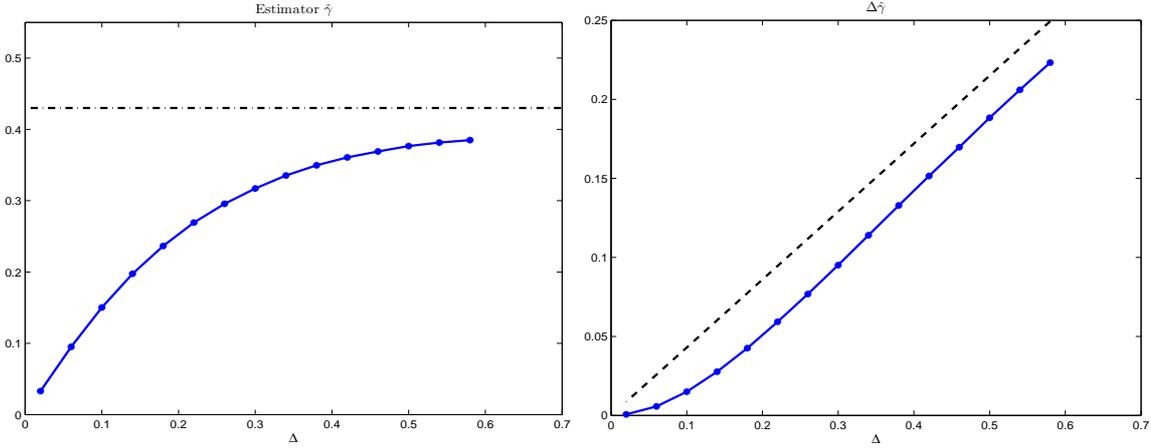
FIG. 5.4. *Left part: estimator $\hat{\gamma}$ vs Δ computed from the TBH dataset with $\epsilon = 0.3$. Solid line: $\hat{\gamma}(\Delta)$ in (2.5), dash-dot line: analytical asymptotic value in (5.15). Right part: $\hat{\gamma}(\Delta)\Delta$ in computed from the TBH dataset with $\epsilon = 0.3$, dashed line: $\gamma\Delta$ where γ is the analytical asymptotic value in (5.15).*
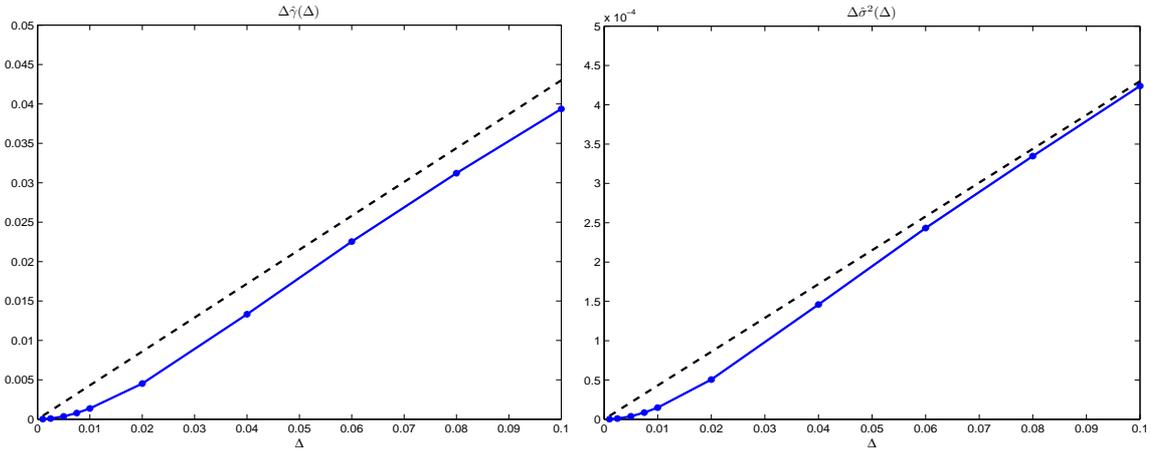
FIG. 5.5. *Behavior of $\hat{\gamma}$ and $\hat{\sigma}^2$ as functions of Δ for data computed from the TBH dataset with $\epsilon = 0.1$. Left part: solid line: $\hat{\gamma}(\Delta)\Delta$ vs Δ, dashed line: straight line with the slope γ in (5.15). Right part: solid line: $\hat{\sigma}^2(\Delta)\Delta$ vs Δ, dashed line: straight line with the slope σ^2 in (5.15). Some points for larger values of Δ are not shown to emphasize the behavior near $\Delta = \epsilon^2 = 0.01$.*

problems of interest, where one needs to study high dimensional and complex multiscale dynamics. When the multi-dimensional diffusion X_t is actually observable, the natural Likelihood function to maximize in order to generate efficient parameter estimators is constructed from a suitable discretization of the SDEs driving X_t. Formally it is then desirable to select the observational time-step as small as possible. On the other hand, the indirect observability context typically imposes a lower bound on the observational time-step and this lower bound depends on the scale parameter ϵ. Moreover, our analytical and numerical results demonstrate that the Maximum
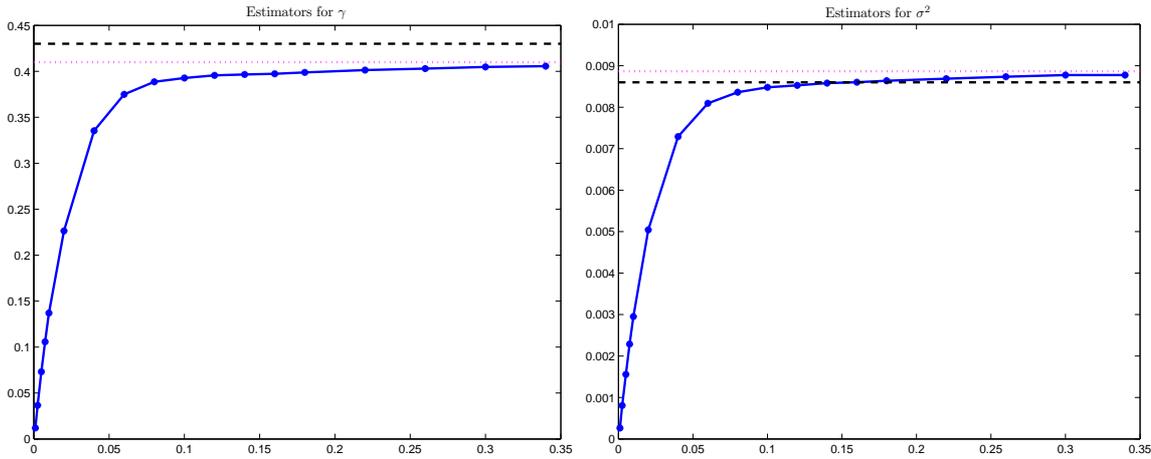
FIG. 5.6. *Estimators as functions of* Δ *computed from the TBH dataset with* $\epsilon = 0.1$. *Left part: solid line:* $\hat{\gamma}(\Delta)$ *in* (2.5), *dotted line:* $\hat{\gamma}_{regression}$ *in* (5.16), *dashed line: analytical asymptotic value in* (5.15). *Right part: solid line:* $\hat{\sigma}^2$ *in* (2.5), *dotted line:* $\hat{\sigma}^2_{regression}$ *in* (5.16), *dashed line: analytical asymptotic value in* (5.15).

Likelihood Estimators computed from an approximating process are biased for any finite value of ϵ. We validated our results on three different multi-scale approximating systems - the triad model, the multiscale potential model, and the TBH model. For all cases studied here, we show that the behavior of correlations of the approximating process Y_t^ϵ near lag zero is crucial for understanding the behavior of the parameters estimators and deriving the appropriate conditions for consistency of parametric estimation. As expected, the time-scale of the fast variables plays the role of the critical scaling which determines the consistency of parametric estimation. These conclusions were verified up to a very high precision.

In this paper we concentrate on studying estimators which are expressed as a differentiable function of lagged covariances. In this case the behavior of covariances is crucial for understanding the behavior of estimators, obtaining sufficient conditions for the consistent estimation, and deriving the speed of convergence of estimators to the "true" values. In particular, we would like to emphasize two main conclusions of our study: (i) sufficient conditions for the consistent estimation and the estimators' speed of convergence can be deduced from the speed of convergence of L^2 or L^∞ norms between the covariances of the approximating and limiting processes and (ii) estimators for different parameters (e.g. the drift vs the diffusion parameter) can depend on different covariances (and their ratio, etc.) and this can result in different speeds of convergence for different parameters. This issue can be particularly important in higher-dimensional systems of equations.

Another important aspect discussed in this paper is estimation of an effective model from a dataset generated with a fixed but unknown value of the scale parameter ϵ. This issue is important in an applied context, since there has been a considerable effort to develop data-driven stochastic parametrization for the large-scale structures from numerical simulations of various geophysical models. Therefore, our results demonstrate that the sub-sampling errors can be the leading cause of discrepancy between the statistical behavior of large scales in full and reduced models and more

elaborate estimation procedures taking the multiscale nature of the data into account should be developed.

We also introduced a regression approach for constructing bias-corrected estimators from a single dataset which is generated by a multi-scale approximating dynamics with a fixed, but unknown, value of the parameter ϵ. This is the realistic parameter estimation context when one is confronted with the task of adjusting a complex dynamic model to physical data for which the correct value of the scale parameter is not directly readable from the data. These results are presented in section 3.2, 4.2, and 5.2 for the triad, multiscale potential, and the TBH model, respectively. To address this problem we treated the sub-sampling time-step Δ as an independent variable and considered parameter estimators as functions of Δ and analyzed their behavior as Δ changes (for a fixed value of ϵ). We showed that the bias due to indirect observability is typically proportional to Δ^{-1} in the favorable sub-sampling regime. We also developed a practical approach for determining the favorable sub-sampling regime by analysis of only the large-scale structures in the data generated by Y_t^ϵ for a fixed but unknown ϵ. We have thus introduced in this context a new class of unbiased parameter estimators derived by analysis (at fixed but unknown ϵ) of a family of standard estimators associated to a whole range of potential sub-sampling time-steps Δ. We showed that the unbiased estimators introduced in this paper can be used as practical tools to compute the adequate range acceptable for the sub-sampling time step Δ, and to practically implement consistent parametric estimation.

The results presented in this paper are also very closely related to recent studies in [32, 31]. In [31] authors considered the sub-sampling problem in a general setting when the observed dynamics is given by a multiscale SDE of the type (2.1). Therefore, the setup is identical to the triad example presented in this paper. In [32] authors consider SDEs with the fast-oscillating potential, including the model in (4.2). We would like to point out that the TBH model is deterministic and, thus, does not fall in either category. In [31] authors rigorously proved (Theorem 4.1 in [31]) consistency of the Maximum Likelihood Estimators under the sub-sampling condition $\Delta = \epsilon^\alpha,\ \ \alpha \in (0,1)$.

On the other hand, our numerical simulations presented in this paper indicate that the sub-sampling criteria on the observational time-step can be extended to cover a larger range of Δ resulting in consistent estimation of parameters. In particular, our criteria is

$$\Delta = \epsilon^{2\alpha},\ \ \alpha \in (0,1). \tag{6.1}$$

For small ϵ, the consistency criteria in (6.1) corresponds to a much large range of acceptable observational time-steps resulting in consistent MLEs. Since MLEs are often constructed using the Euler discretization of the corresponding SDE, smaller observational time-steps are more desirable in practical situations. It is quite possible that the triad example and the model with the fast-oscillating potential presented in this paper fall into a special category, since the limiting process X_t is Gaussian; this issue will be investigated further both analytically and numerically in subsequent papers.

The present study is restricted to realistic parametric estimation of the unobservable Ornstein-Uhlenbeck process, under indirect observability, where the available data set is generated by a specific multiscale dynamics. However, we expect that the methodology presented here is fairly generic and that similar conclusions will hold for other multiscale dynamic models. It is unrealistic to expect that the graph $\hat{\gamma}(\Delta)\Delta$ vs Δ should always be a straight line in the correct sub-sampling regime for all, possibly

nonlinear, effective models. Nevertheless, we expect that the behavior of the MLEs as functions of the variable time-step, Δ, should contain sufficient information to construct the unbiased estimators for the parameters for the underlying model. Future work will include extension of the results presented here to parametric estimation of nonlinear models and systems of SDEs.

Appendix A. Statistical Properties of the Triad Model.

Invariant Density. First, we outline the calculation of the invariant measure for the triad system in (3.1). The Fokker-Planck equation for the invariant measure $P = P(x,y,z)$ for the SDE (3.1) is

$$0 = -\frac{1}{\epsilon}\left[A_1\partial_x(yzP) + A_2\partial_y(xzP) + A_3\partial_z(xyP)\right]$$
$$+\frac{1}{\epsilon^2}\left[s_2\partial_y(yP) + \frac{1}{2}g_2^2\partial_{yy}^2(P) + s_3\partial_z(zP) + \frac{1}{2}s_3^2\partial_{zz}^2(P)\right]. \tag{1.1}$$

The expression in (1.1) may be represented by,

$$0 = \left(\frac{1}{\epsilon}L_{nl} + \frac{1}{\epsilon^2}L_{ou}\right)P,$$

where L_{nl} is the differential operator corresponding to the nonlinear terms and L_{ou} corresponds to the OU terms in the second and third equations in (3.1). The differential operator L_{ou} annihilates the bivariate Gaussian density function given by

$$P_{ou} = K_1\exp\left(-(g_2/s_2^2)y^2 - (g_3/s_3^2)z^2\right).$$

We use the energy conservation assumption by the non-linear terms $(A_1 + A_2 + A_3 = 0)$ and impose that L_{nl} annihilates

$$P = K_2\exp\left(-Bx^2 - (g_2/s_2^2)y^2 - (g_3/s_3^2)z^2\right), \tag{1.2}$$

where B is some positive constant. Substituting (1.2) into (1.1) the constant B becomes

$$B = -A_1^{-1}\left(\frac{A_2g_2}{s_2^2} + \frac{A_3g_3}{s_3^2}\right). \tag{1.3}$$

Therefore, the stationary distribution is given by (1.2) with B given by (1.3), and K_1, K_2 are then the associated normalizing constants. Note, that $Var\{x_t\} = (2B)^{-1} = \sigma^2/(2\gamma) = Var\{X_t\}$, where σ^2 and γ are given by (3.4) and X_t is the limiting Ornstein-Uhlenbeck process in (3.3).

Quadratic Approximation of the Correlation function. To derive the leading-order approximation for the correlation function of x_t we consider a higher order time-discretization scheme

$$x_{t+\Delta} \approx x_t + \frac{A_1y_tz_t}{\epsilon}\Delta + \frac{A_1}{2\epsilon^2}x_t(y_t^2A_3 + z_t^2A_2)\Delta^2$$
$$-\frac{A_1y_tz_t}{2\epsilon^2}(g_2 + g_3)\Delta^2. \tag{1.4}$$

Multiplying (1.4) by x_t and taking expectations with respect to the invariant measure in (1.2) we obtain

$$\mathbb{E}[x_{t+\Delta 1}x_t] \approx \mathbb{E}[x_t^2]\left(1 + \frac{\Delta^2 A_1}{4\epsilon^2}\left(\frac{A_2 s_3^2}{g_3} + \frac{A_3 s_2^2}{g_2}\right)\right),$$

where $\mathbb{E}[x_t^2] = (2B)^{-1}$, with B given by (1.3). Therefore the leading-order approximation for the correlation function $CF_x(\Delta)$ of the slow variable x_t, for $\Delta < \epsilon << 1$, is given by

$$CF_x(\Delta) = 1 - \frac{(g_2+g_3)\gamma}{2\epsilon^2}\Delta^2 + O(\Delta^3), \tag{1.5}$$

where γ, given by (3.4), is the drift coefficient of the reduced SDE. Note that the derivative $(\partial C^\epsilon(\Delta)/\partial\Delta)$ at $\Delta = 0$ is zero.

Appendix B. Computing Variance in the Model with the Fast-Oscillating Potential.

Consider function $f(x) > 0$ which is integrable on R and having the property $\lim_{x\to+\infty} f(x) = \lim_{x\to-\infty} f(x) = 0$. We also assume that $f'(x)$ and $f''(x)$ exist, are continuous and integrable on R. Also, consider a bounded continuous function $g(v) > 0$ on R.

The goal is to compute the behavior of the fast-oscillating integral

$$J(\epsilon) = \int_{-\infty}^{\infty} f(x)g\left(\cos\left(\frac{x}{\epsilon}\right)\right)dx,$$

where the small parameter $\epsilon > 0$ tends to 0.

Let us consider a partition $I_k = [2\pi k\epsilon, 2\pi(k+1)\epsilon]$, $k = -\infty,\ldots,\infty$. Then the integral $J(\epsilon)$ is equal to the infinite sum of integrals over elementary intervals I_k, i.e. $J(\epsilon) = \sum_k J_k(\epsilon)$ where

$$J_k(\epsilon) = \int_{I_k} f(x)g\left(\cos\left(\frac{x}{\epsilon}\right)\right)dx.$$

Since the function $f(x)$ is slowly varying, we use the Taylor expansion to obtain a quadratic approximation for $f(x)$ on each elementary interval I_k. Then an approximation for $J_k(\epsilon)$ can be computed as

$$J_k(\epsilon) = \int_{I_k}[f(2\pi k\epsilon) + (x-2\pi k\epsilon)f'(2\pi k\epsilon) + \frac{1}{2}(x-2\pi k\epsilon)^2 f''(z_k)]g\left(\cos\left(\frac{x}{\epsilon}\right)\right)dx. \tag{2.1}$$

with $z_k \in I_k$.

Next, we can integrate explicitly the fast-scillating function over each elementary interval I_k in (2.1). In particular, if we define the following constants

$$\int_{I_k} g\left(\cos\left(\frac{x}{\epsilon}\right)\right)dx = \epsilon\int_0^{2\pi} g\left(\cos(y)\right)dy = \epsilon Z_0,$$

$$\int_{I_k}(x-2\pi k\epsilon)g\left(\cos\left(\frac{x}{\epsilon}\right)\right)dx = \epsilon^2\int_0^{2\pi} yg\left(\cos(y)\right)dy = \epsilon^2 Z_1,$$

$$\int_{I_k} (x - 2\pi k\epsilon)^2 g\left(\cos\left(\frac{x}{\epsilon}\right)\right) dx = \epsilon^3 \int_0^{2\pi} y^2 g\left(\cos(y)\right) dy = \epsilon^3 Z_2,$$

then the expression (2.1) for $J_k(\epsilon)$ becomes

$$J_k(\epsilon) = Z_0 \epsilon f(2\pi k\epsilon) + Z_1 \epsilon^2 f'(2\pi k\epsilon) + \frac{1}{2} Z_2 \epsilon^3 f''(z_k). \tag{2.2}$$

Substituting (2.2) into the summation for $J(\epsilon)$ we obtain an approximate expression for $J(\epsilon)$

$$2\pi J(\epsilon) = Z_0 \sum_k f(2\pi k\epsilon) 2\pi\epsilon + \epsilon Z_1 \sum_k f'(2\pi k\epsilon) 2\pi\epsilon + \frac{\epsilon^2}{2} Z_2 \sum_k f''(z_k) 2\pi\epsilon,$$

where $z_k \in I_k$ and we also multiplied both sides by 2π. The final step is to treat the infinite summations in the above expression as Riemann sums for the corresponding integrals and obtain the approximation

$$2\pi J(\epsilon) = Z_0 \int_R f(x)\, dx + \epsilon Z_1 \int_R f'(x)\, dx + O(\epsilon^2)$$

where $O(\epsilon^2)$ terms arise due to converting the Riemann sum into the integral and, also, from estimating the remainder term with the second derivative. Since $\int_R f'(x)\, dx = f(+\infty) - f(-\infty) = 0$, we obtain the second order expansion

$$2\pi J(\epsilon) = Z_0 \int_R f(x)\, dx + O(\epsilon^2). \tag{2.3}$$

The second order expansion (2.3) for the integral $J(\epsilon)$ can be used to both, prove the convergence of the variance for the process x_t in the model (4.2) to the variance of the Ornstein-Uhlenbeck process X_t with parameters in (4.3) and, also, to estimate the speed of convergence. Clearly, since the linear term in ϵ is not present in (2.3), the speed of convergence is at least as ϵ^2.

REFERENCES

[1] R. Abramov, G. Kovacic, and A. J. Majda, *Hamiltonian structure and statistically relevant conserved quantities for the truncated Burgers-Hopf equation*, Comm. Pure Appl. Math., 56 (2003), pp. 1–46.

[2] U. Achatz and G. Branstator, *A two-layer model with empirical linear corrections and reduced order for studies of internal climate variability*, J. Atmos. Sci., 56 (1999), pp. 3140–3160.

[3] Y. Ait-Sahalia, P. Mykland, and L. Zhang, *How often to sample a continuous-time process in the presence of market microstructure noise*, Review of Financial Studies, 18 (2005), pp. 351–416.

[4] R. Azencott, A. Beri, P. Ren, and I. Timofeyev, *Sub-sampling in parameter estimation from multiscale data with non-Gaussian limiting behavior.* In preparation, 2013.

[5] R. Azencott, A. Beri, and I. Timofeyev, *Adaptive sub-sampling for parametric estimation of Gaussian Diffusions*, J. Stat. Phys., 139(6) (2010), pp. 1066–1089.

[6] ———, *Parametric estimation of stationary stochastic processes under indirect observability*, J. Stat. Phys., 144(1) (2011), pp. 150–170.

[7] G. Branstator and J. Berner, *Linear and nonlinear signatures in the planetary wave dynamics of an AGCM: phase space tendencies*, J. Atmos. Sci., 62 (2005), pp. 1792–1811.

[8] D. Crommelin and E. Vanden-Eijnden, *Reconstruction of diffusions using spectral data from time-series*, Comm. Math. Sci., 4 (2006), pp. 651–668.

[9] D. Crommelin and E. Vanden-Eijnden, *Diffusion estimation from multiscale data by operator eigenpairs*, Multiscale Model. Simul., 9 (2011), pp. 1588–1623.

[10] T. DelSole, *A fundamental limitation of Markov models*, J. Atmos. Sci., 57 (2000), pp. 2158–2168.

[11] ——, *Stochastic models of Quasigeostrophic turbulence*, Surveys in Geophys., 25(2) (2004), pp. 107–149.

[12] J.-P. Fouque, G. Papanicolaou, R. Sircar, and K. Solna, *Short time-scale in S&P 500 volatility*, J. Comp. Finance, 6 (2003), pp. 1–23.

[13] C. Franzke and A. J. Majda, *Low-order stochastic mode reduction for a prototype atmospheric GCM*, J. Atmos. Sci., 63 (2006), pp. 457–479.

[14] C. Franzke, A. J. Majda, and E. Vanden-Eijnden, *Low-order stochastic mode reduction for a realistic barotropic model climate*, J. Atmos. Sci., 62 (2005), pp. 1722–1745.

[15] M. Hoffmann, *Rate of convergence for parametric estimation in a stochastic volatility model*, Stochastic processes and their applications, 97 (2002), pp. 147–170.

[16] I. Horenko, E. Dittmer, A. Fischer, and C. Schutte, *Automated model reduction for complex systems exhibiting metastability*. To appear in Multiscale Model. Simul., 2006.

[17] G. Hummer, *Position-dependent diffusion coefficients and free energies from Bayesian analysis of equilibrium and replica molecular dynamics simulations*, New Journal of Physics, 7 (2005), p. 34.

[18] R. Z. Khasminsky, *A limit theorem for the solutions of differential equations with random right-hand sides*, Theory Prob. Applications, 11 (1966), pp. 390–406.

[19] ——, *On stochastic processes defined by differential equations with a small parameter*, Theory Prob. Applications, 11 (1966), pp. 211–228.

[20] S. Kravtsov, D. Kondrashov, and M. Ghil, *Multilevel regression modeling of nonlinear processes: Derivation and applications to climatic variability*, Journal of Climate, 18(21) (2005), pp. 4404–4424.

[21] T. G. Kurtz, *A limit theorem for perturbed operator semigroups with applications to random evolution*, J. Funct. Anal., 12 (1973), pp. 55–67.

[22] A. Majda, C. Franzke, and D. Crommelin, *Normal forms for reduced stochastic climate models*, Proc. Nat. Acad. Sci. USA, 16(10) (2009), pp. 3649–3653.

[23] A. J. Majda and I. Timofeyev, *Remarkable statistical behavior for truncated Burgers-Hopf dynamics*, Proc. Nat. Acad. Sci. USA, 97 (2000), pp. 12413–12417.

[24] ——, *Statistical mechanics for truncations of the Burgers-Hopf equation: a model for intrinsic stochastic behavior with scaling*, Milan Journal of Mathematics, 70(1) (2002), pp. 39–96.

[25] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden, *Models for stochastic climate prediction*, Proc. Nat. Acad. Sci. USA, 96 (1999), pp. 14687–14691.

[26] ——, *A mathematics framework for stochastic climate models*, Comm. Pure Appl. Math., 54 (2001), pp. 891–974.

[27] ——, *A priori tests of a stochastic mode reduction strategy*, Physica D, 170 (2002), pp. 206–252.

[28] ——, *Systematic strategies for stochastic mode reduction in climate*, J. Atmos. Sci., 60(14) (2003), pp. 1705–1722.

[29] ——, *Stochastic models for selected slow variables in large deterministic systems*, Nonlinearity, 19(4) (2006), pp. 769–794.

[30] G. Papanicolaou, *Some probabilistic problems and methods in singular perturbations*, Rocky Mountain J. Math, 6 (1976), pp. 653–673.

[31] A. Papavasiliou, G. A. Pavliotis, and A. Stuart, *Maximum likelihood drift estimation for multiscale diffusions*, Stoch. Proc. and Applics., 119(10) (2009), pp. 3173–3210.

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

[33] G. A. Pavliotis and A. M. Stuart, *Multiscale Methods: Averaging and Homogenization*, Springer, 2008.

[34] C. Schutte, J. Walter, C. Hartmann, and W. Huisinga, *An averaging principle for fast degrees of freedom exhibiting long-term correlations*, Multiscale Model. Simul., 2(3) (2004), pp. 501–526.

[35] P. Sura and J. Barsugli, *A note on estimating drift and diffusion parameters from timeseries*, Phys. Lett. A, 305 (2002), pp. 304–311.

[36] L. Zhang, P. Mykland, and Y. Ait-Sahalia, *A tale of two time scales*, Journal of the American Statistical Association, 100 (2005), pp. 1394–1411.