ESTIMATION OF STOCHASTIC MODELS UNDER INDIRECT OBSERVABILITY

A Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Arjun Beri

August 2010

ESTIMATION OF STOCHASTIC MODELS UNDER INDIRECT OBSERVABILITY

Arjun Beri

APPROVED:

Dr. Ilya Timofeyev (Chairman)
Department of Mathematics, University of Houston

Dr. Robert Azencott (Co-Chairman)
Department of Mathematics, University of Houston
Emeritus Professor, ENS, France

Dr. Bernhard Bodmann
Department of Mathematics, University of Houston

Dr. Gregor Kovačič
Department of Mathematics, RPI, Troy, New York

Dr. John L. Bear
Dean, College of Natural Sciences and Mathematics,
University of Houston

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to both my advisors, Dr. Ilya Timofeyev and Dr. Robert Azencott, who have patiently guided me through to bring this work to fruition. Their constant encouragement, active guidance throughout my research, and their profound insights in the subject, which they shared with me, have been the pillars for the successful completion of this difficult, yet exceedingly fruitful, task. I have been extremely fortunate to have studied mathematics under their supervision. I also acknowledge the generous financial support that I received as a research assistant from both Dr. Robert Azencott as well as Dr. Ilya Timofeyev, during the last three years. I am also thankful to Department of Mathematics at University of Houston for providing me with ample funding and an excellent environment to pursue my graduate studies.

I sincerely thank my committee members Dr. Bernhard Bodmann and Dr. Gregor Kovačič for devoting their valuable time towards reading my thesis, and making very useful suggestions.

Various faculty members in the Department of Mathematics have made contributions to my life as a graduate student in many ways. I thank Dr. Alexandre Caboussat, Dr. Edward Kao, Dr. Andrew Török, Dr. Mathew Nicol, Dr. Shanyu Ji, and graduate advisor Ms. Pamela Draughn, for their guidance.

I also thank my friends, colleagues at University of Houston, especially, Yutheeka Gadhyan, Dr. Kawin Nimsaila, and Dr. Nofil Barlas, for discussions about mathematics and everything else under the sun.

There are numerous people who are "behind the scenes" in this work, namely, my sister Aditi, and brother-in-law Ashutosh, who have showered their love and affection on me. Lastly, the contribution of my parents, father Dr. Rakesh Narain Beri, MD, and mother Romi Beri, towards my achievements is immeasurable. I thank them to allow me to pursue my ambitions, and for instilling in me the fearless spirit of exploration and enquiry. I will

forever remain indebted to them.

I dedicate this thesis to the memory of my dear father Dr. Rakesh Narain Beri. His grasp of scientific knowledge, philosophy, his command over languages, his elegant Tennis strokes, and his goodness, will always inspire me in all my endeavors.

ESTIMATION OF STOCHASTIC MODELS UNDER INDIRECT OBSERVABILITY

An Abstract of a Dissertation

Presented to

the Faculty of the Department of Mathematics

University of Houston

In Partial Fulfillment

of the Requirements for the Degree

Doctor of Philosophy

By

Arjun Beri

August 2010

ABSTRACT

Statistical estimation of parameters in a carefully chosen stochastic model given a discrete dataset is an important requirement in various scientific studies. The available data are not generated from the stochastic model under investigation, but are typically observed in nature, or come from the financial markets, or are numerically simulated from complex dynamics. Therefore, there may be inherent differences between the data and the model. For instance, they may not agree at different time scales, which renders estimation of parameters difficult.

The thesis is presented in three parts. In part I (chapters 2, 3) of the thesis we introduce the parametric estimation set-up by considering a particular example of a stochastic differential equation, namely, the Ornstein-Uhlenbeck process. The estimators for the drift and diffusion coefficients are introduced with a detailed study of their convergence properties, first in the benchmark case, where the *direct* observations from the underlying Ornstein-Uhlenbeck process are assumed to be available. Most results pertaining to this *direct observability* case are classical, but are introduced as a benchmark for comparison with the case of estimation under *indirect observability*. *Indirect observability* framework refers to the scenario in which direct observations of the process that is being estimated are unavailable, and the available observations are sampled from only an approximating process. *Indirect observability* framework is, then, introduced with complete characterization of the optimal subsampling schemes that guarantee *asymptotic consistency* of the parameter estimators, in particular, for a specific approximating process named the *smoothed Ornstein-Uhlenbeck* process. The results obtained from the study of this specific process forms the backbone of the subsequent analysis.

In part II (chapters 4, 5), validity of the derived optimal subsampling schemes is verified numerically in a more complex setting, namely, parametric estimation of a limiting stochastic model given observations from the associated full multiscale system. Estimation of these

reduced models naturally falls in the *indirect observability* framework, as defined here. Moreover, the numerical results, along with the theoretical arguments, help to draw similarities in the asymptotic behavior of estimators based on the relatively complex multiscale dynamics, with those from the simple smoothed Ornstein-Uhlenbeck process. Chapter 5 presents a novel methodology to estimate parameters under indirect observability in a practical setting, i.e., given large but finite number of discrete observations from an approximating process. Specifically, using the values of the estimators based on multiple datasets obtained by subsampling the given single dataset, the methodology introduces bias-corrected estimators, and also computes a precise estimate for the implict degree of approximation (or scale-separation) of the dataset as compared to the *true unobserved* process.

Part III (chapter 6) extends *indirect observability* to a general mathematical framework to study estimation of parameters associated to an *unobserved* centered stationary Gaussian process. The *observed* approximating process is a centered stationary process with arbitrary probability distribution.

Contents

List of Figures

Chapter 1

Background and Motivation

A story should have a beginning, a middle, and an end ... but not necessarily in that order.
– Jean-Luc Godard

1.1 Introduction

Elements of randomness have long been introduced, via stochastic modeling, in representing large-scale deterministic systems by simplified equations, and to study their properties only probabilistically. Also, in many practical problems it is desirable to model the dynamical features of observed datasets by carefully chosen stochastic models. Then, a valid objective is to estimate the model parameters to best-fit the given observations. This is the central theme of the thesis. However, the motivation to consider the estimation problem in this study comes from a common pathological situation, unobservability. Often the data to be fitted are only an approximation to the model. Moreover, the experimental or observational datasets in nature (or financial markets) are not explicitly generated by the underlying stochastic model and are, thus, expected to agree with the model only in coarse statistical sense. Hence, parametric estimation of stochastic models from such datasets may not be robust. In such scenarios one must show caution in applying the standard statistical

techniques to obtain optimal estimates for the model parameters such as maximum likelihood estimation. Such behavior is well documented and widely studied; for instance, in the context of market microstructure noise in high frequency financial timeseries [3, 2, 12, 5], in molecular dynamics [28, 48], and in ocean-atmosphere sciences [15, 27, 34] where the data exhibit multiscale character.

The multiscale nature of complex dynamics has been a particularly active area of research in the last few decades. 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 [38, 47, 46, 86]. Long-term evolution of high-dimensional deterministic systems governed by complex PDEs, have often been approximated by low-dimensional reduced stochastic models focused on larger time scales and with a good statistical fit to the observed dynamic data. For instance, the stochastic mode-reduction technique [63, 64, 66, 65] has successfully modeled the dynamics of large-scale structures in systems with time-scale separation, an optimal prediction setup has enabled coarse-grained dynamic modeling of statistical descriptors [22, 20, 21]; spin-flip processes have provided coarse-grained models of traffic flow [53, 54, 55, 4]; reduced Markov chain models have been applied to prototype atmosphere-ocean interactions [25, 68]; and a generic framework has been developed for dimension reduction in metastable systems [82, 45]. In most practical contexts of this type, one seeks to approximate the dynamics of key statistical descriptors of a chaotic high-dimensional deterministic dynamics by a closed form low-dimensional stochastic process, such as (vector) stochastic differential equations (denoted as SDE). Then the available data $U_n = Y_{n\Delta}$, $n = 1, 2, ...N$ are not generated by the underlying SDE, but sampled from observations Y_t generated by some complex, not completely identifiable, deterministic dynamics. At short time scales, the trajectories Y_t of the observable physical

process are quite different from sample paths X_t of a vector SDE (see for instance [27]), but on longer time scales the behavior of Y_t is well emulated by X_t. This situation is typical for data generated by a numerical dynamic model, such as fluid dynamics PDEs. The Y_t trajectories then decorrelate slower than those of X_t, and good estimators $f(X_{t1}...X_{tN})$ of the underlying SDE parameters can lose their consistency if one simply substitutes X_t for Y_t in the function f and uses observations $(Y_{t1}...Y_{tN})$ which are too dense in time.

A typical phenomenon empirically observed in practice is the existence of high-frequency observational noise in a time series. In various applications, for instance, stochastic modeling in finance [3, 2, 12, 5, 39, 44] and oceanography [41], the dataset is contaminated by high frequency observational noise. Estimation studies under high frequency observational noise assume that the observed discrete datasets are decomposable into an unobservable true value X_t, and noise term $\mathcal{E}_i$, namely,

$$Y_{t_i} = X_{t_i} + \mathcal{E}_i,$$

where noise $\mathcal{E}_i$ has standard deviation denoted by $\sigma_e > 0$. Under such assumptions the realized variance (for example, [39, 44]), otherwise well behaved estimator of the integrated volatility process associated to X_t, is no longer consistent when computed from the contaminated observations Y_{t_i}. Estimation of the underlying model based on datasets with multiscale character and for data contaminated with high-frequency observational noise have common features. For instance, in both cases the model is not compatible with the data at all time scales. Also, in both situations considering discrete observations on a more coarse time-grid has been a proposed recipe to obtain accurate estimates.

The thesis proposes a formal mathematical framework to undertake parametric estimation of stochastic models with "contaminated" observations. This *indirect observability* is mathematically formulated in the following generic manner. Consider discretely sampled data $U_n = Y^{\varepsilon}_{n\Delta}$, $n = 0, 1, 2, ...N$, from an observable process Y^{ϵ}_t such that $Y^{\varepsilon}_t \to X_t$, in

some adequate sense, as $\varepsilon \to 0$. Hence, Y_t^ε is an approximation of the underlying process $X_t = X_t(\theta)$, where parameter vector $\theta \in \Theta$, such that Θ is an open bounded subset of $\mathbb{R}^p$. In such a scenario the statistical properties of the estimators $\hat{\theta}_\varepsilon$ of θ based on the discrete observations of the approximating process Y_t^ε do not follow naturally. This study provides explicit conditions on the number of discrete observations $N = N(\varepsilon)$, and the uniform time-step $\Delta = \Delta(\varepsilon)$ such that, as $\varepsilon \to 0$, the estimator $\hat{\theta}_\varepsilon$ based on the observations from the approximating process Y_t^ε converges to the true parameter value θ. This idea will be made concrete and formal in the ensuing chapters.

1.2 Precursor

The above objective is presented step-by-step in the following manner.

1. Estimation under *indirect observability* framework is first presented for a specific model, namely, the Ornstein-Uhlenbeck (OU) process. The OU process is given as an exact solution of a linear SDE whose dynamics are given by,

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

   where $\gamma > 0,\ \sigma > 0$ are the parameters characterizing the process. It is assumed that although the parameters γ, σ are to be estimated, the process X_t can not be observed.

2. The parameters are estimated based on observations from an approximating process Y_t^ε, which is defined to be the local average of the OU process over a small moving window of time, i.e.,

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

   The process Y_t^ε is called the *Smoothed Ornstein-Uhlenbeck* process (denoted as SOU).

3. This specific pair of processes (X_t, Y_t^ε) provides us with an excellent prototypical

model to study *indirect observability* mainly because of the Gaussianity (See chapters 2, 3) of the processes. In particular, with Gaussian density known in closed-form, exact computations of asymptotic expressions for the L_2-errors in estimators for γ, σ are possible.

4. Indirect observability framework of the OU/SOU processes provides an exact indication of how the observations sampled from the approximating process Y_t^ε impact the estimators. These results when extended to estimation of reduced (c.f., homogenized) equations given observations from large-dimensional complex systems with multiple time scales, provide remarkable tenacity. In particular, we present detailed numerical study of two multiscale dynamics - the additive triad model [7, 9, 64], and the truncated Burgers-Hopf model [1, 61, 62].

5. A novel methodology is then presented to estimate parameters under indirect observability in a practical setting, i.e., given large but finite number of discrete observations from an approximating process with a fixed unknown value of the small parameter ε. Specifically, using the values of the estimators based on multiple datasets obtained by subsampling the given single dataset, the methodology introduces bias-corrected estimators, and also computes a precise estimate for the implicit degree of approximation, ε, (or scale-separation) of the dataset as compared to the *true unobserved* process. This numerical technique is verified in the case of the OU/SOU example, and with the two multiscale dynamic models.

6. In the last part of the thesis we extend the methodology to a fairly general class of models. The unobservable process X_t is assumed to be a centered stationary Gaussian process, and the unobservable process Y_t^ε is allowed to have arbitrary distributions, as opposed to strict Gaussianity case of the OU/SOU process. The observable process X_t is assumed to have an associated covariance function $K(u) = K(u, \theta)$, parametrized

by vector θ. An objective, then, is to consistently estimate θ, as $\varepsilon \to 0$, by observing the process Y_t^ε, converging to X_t in an adequate sense.

7. With this extension to the general framework, we also present the notion of estimators with *non-vanishing lags*, which proves to be extremely useful in the indirect observability scenario. These estimators, characterized by the asymptotic conditions on N, Δ and the discrete lags k (precisely defined below); *do not see the small scales* thereby rectifying the issue of multiple time scales, and are an improvement to the naive subsampling of data.

The parametric approach is often utilized for the purpose of stochastic modeling. Parameters characterizing the stochastic model need to be identified from the set of observational or experimental data. In particular, estimating parameters of a system of stochastic differential equations (SDEs) from a discrete dataset of observations has been the center of applied research for many decades [78, 83, 17]. On the other hand, non-parametric estimation of stochastic models do not assume any fixed parametrization, and is used under various scenarios. Various non-parametric techniques for estimating the drift and diffusion terms from discrete data (see, for example, [13, 24]) can be used in low-dimensional systems. Nevertheless, due to the data limitations these techniques are applicable only if the dimension of the SDE is less than four. Therefore, here we consider the more general parametric approach for estimating the underlying models given the "contaminated" data.

Some of the notation and terminology used in the thesis is as follows. The process of identifying the *optimal* time interval (or, the frequency of sampling) between observations that leads to *asymptotically consistent estimators*, is referred to as *optimal subsampling criterion*. The framework under which the parametric estimation of the underlying unobservable process is based on the data from an observable approximating process, is termed as *indirect observability* framework. On the other hand, if the process X_t is observable, then,

the parametric estimation is said to be under *direct observability*, which is the benchmark scenario.

In general, estimation is based on a set of discretely subsampled observations $U_n = U_n(\Delta)$, with uniform time step $\Delta > 0$, from a given observable continuous-time random process. The estimators are functions of the discrete observations, the number of observations N, and the time interval Δ, namely, $\hat{\theta} = \hat{\theta}(N, \Delta, \{U_n; n = 0, \ldots, N-1\})$. Under the classical direct observability framework, $U_n = X_{n\Delta}$; and under indirect observability the discrete observations $U_n = Y^{\varepsilon}_{n\Delta}$. The asymptotic properties of the estimator $\hat{\theta}$ are an indicator of the efficacy of the estimation procedure and reliability of the estimated model. Therefore, a minimum requirement is that as $N \to \infty$ the estimator $\hat{\theta} \to \theta$, where the convergence to the true value θ is, *at least, in probability*, or where possible it must be *convergence in* L^2. An estimator $\hat{\theta}$ satisfying this property is said to be *asymptotically consistent*.

In the asymptotic study of the estimators, under indirect observability, considered for the parameters in vector θ, or in particular, the estimators for the OU parameters γ, σ^2, we consider the following equivalent design of estimation. Assume that estimation is based on $N = N(\varepsilon)$ discrete observations sampled from the observed process with a uniform time step $\Delta = \Delta(\varepsilon) > 0$. Then, as the small parameter $\varepsilon \to 0$, one or several adaptive subsampling schemes are applied to observed process which, specifically under *indirect observability* is Y^{ε}_t; and under *direct observability* is the true process X_t. These adaptive subsampling schemes are identified by the following conditions,

$$\varepsilon \to 0, \qquad \Delta = \Delta(\varepsilon) \to 0, \qquad N = N(\varepsilon) \to \infty \ \text{ such that } \ N\Delta \to \infty.$$

Under indirect observability as $N \to \infty$, with ε and Δ fixed, the estimators are inconsistent with a non-vanishing bias. However, the asymptotic behavior of estimators may still provide important clues to the *optimal subsampling criterion*. In practice, the estimation is based

on one large dataset subsampled from a single trajectory generated by complex systems with a fixed unknown value of ε. Then in such situations to understand the limitations of the estimation process it is crucial to obtain partial asymptotics with just $N \to \infty$, which is referred to as *fixed-rate subsampling schemes*. Under direct observability, the maximum likelihood estimators, in particular, for the OU parameters are asymptotically consistent given the fixed-rate subsampling scheme.

Most asymptotic results for the parametric estimation of Gaussian models, in particular, the OU process, are classical under the *direct observability* scenario. Nevertheless, these results are described in chapter 2 with an emphasis on the behavior of the estimators under the adaptive subsampling scheme. The approach presented in chapter 2 is carried over to the case of the indirect observability case. It is showed that there is a specific subsampling criterion for N and Δ that guarantees asymptotic consistency of the estimators under direct observability. As expected, these adaptive subsampling conditions are not preserved under indirect observability, and stronger conditions are required for asymptotic consistency.

The focus of this thesis is on the situation when the underlying true process is unobserved and the parameters are estimated based on the observed *approximating process*. In chapter 3, for brevity, indirect observability is first illustrated for estimation of the drift and diffusion parameters γ, σ^2, in the stochastic differential equation corresponding to the Ornstein-Uhlenbeck (OU) process, namely,

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

OU process is the stochastic model of choice in various applications [50, 42, 60]. In our study this specific example provides rigorous study of the issue of indirect observability and subsampling. It is assumed that the observable approximating process is the *smoothed Ornstein-Uhlenbeck (SOU) process* which is obtained from averaging the Ornstein-

Uhlenbeck process over a small moving window ε of time. These two processes are stationary and Gaussian with known densities, which makes their study exact. Therefore, in this particular case the characterization of the *optimal subsampling criterion* is plausible. As shown below, the criterion identified for the SOU/OU example holds true for other more complex examples, for instance, the additive triad system and the truncated Burgers-Hopf system, as presented in chapter 4. The framework for the parametric estimation of OU process based on observations from SOU process has been presented in [7].

In chapter 5, a new methodology to accurately estimate parameters is presented which is applicable in practical situations, that is, when the available discrete observations come from a single trajectory that corresponds to approximating dynamics with a *fixed unknown* ε. In such a setting, the otherwise well characterized adaptive subsampling schemes in terms of the small parameter ε are no longer useful. Hence, an alternate view of the problem is considered and estimators derived from multiple datasets, by subsampling the given dataset, are used to estimate the unknown ε, and infer the critical optimal regime to accurately estimate the parameters. The results of chapter 5 are based on [9].

Chapter 6 presents an extension of the indirect observability framework to a more general class of processes. A centered stationary Gaussian process X_t is considered as the *unobservable process* that needs to be identified given discretely sampled observations. A zero-mean Gaussian process is completely characterized by its second-order moments, or the covariance function. In this general situation the covariance function $K(u) = E[X_t X_{t+u}]$ associated to the process X_t is assumed to be parametrized by an unknown vector θ of parameters. With some regularity conditions on the decay rates, and differentiability of the covariance function, the estimation of the parameter θ is undertaken based on an observable approximating process Y_t^ε. There are few assumptions on the distribution of the observable approximating process Y_t^ε, that it is uniformly bounded in L_4-norm, and $Y_t^\varepsilon \to X_t$, as $\varepsilon \to 0$, in L_4, such that $\|Y_t^\varepsilon - X_t\|_{L_4} \leq \rho(\varepsilon) \to 0$. Then, the adaptive subsampling schemes

ensuring the asymptotic consistency of the estimators, are well defined and can be characterized by $\rho = \rho(\varepsilon)$, number of observations $N(\varepsilon)$, and uniform time-step $\Delta(\varepsilon)$. The results of chapter 6 are based on [8].

1.3 Estimation of Stochastic Differential Equations

Consider a *filtered probability space*, i.e., a *complete probability space* $(\Omega, \mathcal{F}, P)$ together with a *filtration* $\{\mathcal{F}_t\}$ satisfying the *usual hypothesis*. This means $(\Omega, \mathcal{F})$ is a measurable space, and P is a probability measure on $(\Omega, \mathcal{F})$ such that $\mathcal{F}$ contains all subsets of each $P-$null set in $\mathcal{F}$. A *filtration* $\{\mathcal{F}_t\}$ is a family of sub-σ-algebras of $\mathcal{F}$ such that $\mathcal{F}_s \subset \mathcal{F}_t$, for all positive real numbers $s < t$. The filtration is said to satisfy the *usual hypothesis*, if the following conditions hold,

1. Right continuity : $\mathcal{F}_t = \bigwedge_{s>t} \mathcal{F}_s$, for all t,
2. Completeness : $\mathcal{F}_0$ contains all $P-$null sets in $\mathcal{F}$.

We define all the stochastic processes of interest on this probability space. A one-dimensional stochastic differential equation (see [70, 79, 23]), denoted by SDE, is given by

$$dX_t = a(t, X_t)dt + b(t, X_t)dW_t, \qquad \text{for } \ t \geq 0, \tag{1.1}$$

where W_t is a standard Brownian motion [49] (see appendix 0.12), and the initial condition X_0 is such that $E[|X_0|^2] < \infty$. The terms $a(t, X_t)$, $b(t, X_t)$ are scalar functions, respectively, from $(\mathbb{R}_+ \times \mathbb{R}) \to \mathbb{R}$. The variable t is considered to represent the evolution of time and X_t is the state of the system. The existence and uniqueness of the solution to (1.1) is a consequence of the regularity conditions (Lipschitz and bounded growth) on the coefficient functions $a(t, X_t)$ and $b(t, X_t)$. The differential equation in (1.1) is only a notation, the SDE is interpreted as an integral equation where the second term is an Itô's integral. The

term $a(t, X_t)$ is referred to as the drift function, and $b(t, X_t)$ is the diffusion function. The drift and diffusion coefficient functions, respectively, are interpreted as instantaneous mean and variance, namely,

$$a(t, X_t) = \lim_{h \to 0} E\{h^{-1}(X_{t+h} - X_t) \mid X_t(\omega) = X_t\},$$

$$b^2(t, X_t) = \lim_{h \to 0} E\{h^{-1}(X_{t+h} - X_t)^2 \mid X_t(\omega) = X_t\}.$$

The detailed description of the SDE is given in, for instance, [6, 70, 79, 23].

Continuous-time diffusion processes are often used to model physical systems in nature, and financial markets, for instance, the stock prices, interest rates, or currency exchange rates. In practice, however, the diffusions are not observed continuously, but only at discrete time points. Many applications involve estimation of the drift and diffusion coefficients given a stream of discrete observations U_n.

Broadly, there are two approaches to compute the coefficients $a(t, X_t)$ and $b(t, X_t)$ given discrete datasets. First, referred to as *parametric estimation*, is the technique where one assumes a specific parametrization of the drift and diffusion coefficients, namely, a vector θ such that $a(t, X_t) = a(t, X_t; \theta)$, and $b(t, X_t) = b(t, X_t; \theta)$, where the unknown parameter vector $\theta \in \Theta \subset \mathbb{R}^p$ for some open bounded parameter space Θ. The second approach is the *non-parametric estimation* where no a priori functional form is specified, and the coefficients are estimated locally in a neighborhood of every spatial point that the process visits (see for instance [11, 26, 51, 13]).

For brevity, we will focus only on parametric estimation to emphasize on indirect observability framework, and illustrate the impact of indirect observability on estimation. The main goal of this study is to present parametric estimation of a stochastic process $X_t(\theta)$ when the available observations are subsampled from an approximating process $Y_t^{\varepsilon}(\theta)$.

The method of maximum likelihood estimation is used to obtain estimators for the unknown vector of parameters θ. However, estimation of generic nonlinear diffusion processes based on the exact likelihood function is possible only in very special cases, precisely when the solution, and hence, the transition density; to the SDE is available. Therefore, in general the transition density (or the probability distribution) is not known in closed form. Various approaches have been considered in literature to approximate the likelihood function, or to devise other semi-parametric techniques to obtain estimators(see for instance, [67] for an overview). We start our presentation with the initial objective of undertaking the estimation study by considering a widely used approach to construct an estimator, and study the impact of indirect observability on it. However, in the course of our study we also show that the estimators may be modified to better suit the framework of indirect observability, as seen in chapter 6, where we present the extended framework, and introduce the generic notion of estimators with *non-vanishing lags*.

Next, we outline a particular approach, based on the Euler-Maruyama discretization of the SDE, to obtain the approximate likelihood function for the diffusion process. Then, using this approximate likelihood function we obtain the so called quasi-maximum likelihood estimators for θ.

Parametric Estimation of SDEs

In our study we carry out parametric estimation of the stochastic differential equations. We assume that we are given a set of discretely subsampled observations U_n, $n = 0, 1, \ldots N$, such that $U_n = X_{t_n}$, for the time points of observations $t_0 < t_1 < \ldots < t_N$. Conditional on this set of observations we compute a statistical estimate $\hat{\theta}$ for the parameter θ in

$$dX_t = a(t, X_t; \theta)dt + b(t, X_t; \theta)dW_t. \tag{1.2}$$

We use the statistically efficient, Maximum Likelihood technique to estimate the un-

known parameter θ. However, in many cases the transition density of X_t, given by (1.2), is not known in a closed form. In such situations, one approach [78] may be to compute an approximation to the likelihood function based on the discretized version of the SDE. In particular, given initial value $\tilde{X}_{t_0} = X_{t_0}$, we may use the Euler-Maruyama (denoted as EM) discretization scheme to approximate the solution X_t of (1.2) over a small interval $[t_0, t_0 + h]$ by

$$\tilde{X}_{t_0+h} = \tilde{X}_{t_0} + a(t_0, \tilde{X}_{t_0}; \theta)h + b(t_0, \tilde{X}_{t_0}; \theta)(W_{t_0+h} - W_{t_0}) + O(h^{1/2}). \tag{1.3}$$

When the exact solution X_t is not known we may assume that the observations U_n satisfy (1.3). Since the increments of the Brownian motion are Gaussian, therefore, (1.3) provides us with a Gaussian approximation (with first two moments equivalent to those of the process X_t) to the exact conditional transition density function associated to X_t. The convergence properties of the discretization schemes for the SDEs are presented in [58]. One may often use the higher order Milstein scheme but here we utilize the Euler-Maruyama scheme since the gain in the precision is not commensurate with the increase in complexity. The likelihood based estimators computed using the stochastic difference equations (1.3) are in general inconsistent, and biased for a fixed sampling time step. We systematically consider the adaptive subsampling schemes with time-step $h = \Delta \to 0$, therefore, the estimators based on the above discretization are justified in our study.

Assume a fixed uniform sampling time step $\Delta > 0$ for the observations, then, $U_n = X_{n\Delta}$, for $n = 0, 1, \ldots, N-1$, satisfy

$$U_{n+1} \approx U_n + a(t_n, U_n; \theta)\Delta + b(t_n, U_n; \theta)(W_{t_{n+1}} - W_{t_n}), \tag{1.4}$$

where the initial value U_0 is a constant in $\mathbb{R}$. Therefore, the log likelihood function is

$$\mathbb{L}_N(\theta) = -\frac{1}{2}\sum_{n=0}^{N-1}\left(\log\left(2\pi\Delta b^2(t_n, U_n;\theta)\right) + \frac{1}{\Delta}\left(\frac{U_{n+1} - U_n - a(t_n, U_n;\theta)\Delta}{b(t_n, U_n;\theta)}\right)^2\right). \quad (1.5)$$

The principle of maximum likelihood is to find an estimate $\hat{\theta}$ in a compact subset of Θ that maximizes the likelihood of the given trajectory identified by the sample U_n. The first-order necessary condition are such that the *score function* $\nabla\mathbb{L}_N(\theta) = 0$. This determines a system of nonlinear equations in θ whose solution provides an expression for the estimators. We assume that the coefficients are such that the gradient vector $\nabla\mathbb{L}_N(\theta)$ exists in a neighborhood of the true value. It is not always possible to compute a closed-form solution to the system of equations identified by the necessary conditions $\nabla\mathbb{L}_N(\theta) = 0$, therefore, numerical optimization techniques will be used, for instance, the gradient method [16, 69], to obtain the estimates.

Estimation of a Stationary Gaussian Process

In the general framework presented in the chapter 6, *indirect observability* is studied under the assumption that the observable process is a centered stationary Gaussian process X_t. The covariance function associated to the process X_t is given as $K(u) = K(u, \theta),\ \forall u \in \mathbb{R}$, and the unknown parameter vector $\theta \in \Theta$, where Θ is a bounded open subset of $\mathbb{R}^p$. With no specific functional form assumed for the covariance function, except that it decays exponentially for $|u| \to \infty$, the application of the likelihood approach is not feasible. However, we apply a particular version of the method of moments to derive estimator for the vector θ.

Since centered Gaussian processes with the same covariance functions are probabilistically equivalent, it is quite natural to assume that the covariance function $u \to K(u, \theta)$ determines θ uniquely. A slightly stronger hypothesis is to assume the existence of a finite fixed

set of time lags $\{u_1, ..., u_p\}$ such that the vector $r = (r_1, \ldots, r_p)$ of covariances $r_j = K(u_j, \theta)$ determines a unique $\theta \in \Theta$ denoted by $\theta = F(r)$. Then, for each $j = 1, \ldots, p$, consider the covariance estimators $\hat{r}_j$ based on discrete observations of the underlying observed process, then the estimator for the vector θ is given by $\hat{\theta} = F(\hat{r})$, where $r = (\hat{r}_1, \ldots, \hat{r}_p)$.

Asymptotic Properties of the Estimators

Some of the properties of optimal estimators are defined below.

Definition 1.3.1 (Consistency) *A statistical estimator $\hat{\theta}(N)$, based on N discrete observations (U_n) is said to be consistent, $\iff$ $\hat{\theta}(N) \to \theta_0$ in probability, as $N \to \infty$. In some situations, for instance Gaussian processes, the convergence may hold in the sense of L^2.*

Definition 1.3.2 (Gaussianity) *A statistical estimator $\hat{\theta}(N)$, based on N discrete observations (U_n) is said to be Asymptotically Gaussian $\iff$ as $N \to \infty$, the random variable $s(N)(\hat{\theta}(N) - \theta) \to \mathbf{N}(0, \Sigma)$, where $s(N)$ is the normalizing factor and Σ is the asymptotic covariance matrix.*

Definition 1.3.3 (Efficieny) *A statistical estimator $\hat{\theta}(N)$, based on N discrete observations (U_n) is said to be efficient $\iff$ $var(\hat{\theta}(N))$ attains the Cramér-Rao bound [80], for all values of the estimator θ. Some estimators may attain this bound only asymptotically and are said to be asymptotically efficient.*

In our study, the optimal estimators are subjected to extended asymptotics, namely, the systematic conditions under the adaptive subsampling scheme, i.e., as $\varepsilon \to 0$, $\Delta = \Delta(\varepsilon) \to 0$, $N(\varepsilon)\Delta(\varepsilon) \to \infty$. Therefore, the above asymptotic properties are considered under the adaptive subsampling schemes as $\varepsilon \to 0$.

Lastly, corresponding to a discrete centered stationary process $U = \{U_n\}_{n\in\mathbb{N}}$, define the standard *empirical covariance estimators* $\hat{r}_k(N)$, for each $k \geq 0$, as

$$\hat{r}_k(N) = \frac{1}{N}\sum_{n=0}^{N-1} U_{n+k}U_n. \tag{1.6}$$

These estimators are asymptotically consistent, efficient, and asymptotically Gaussian (see [10, Chapter X]) for a centered stationary *Gaussian* process U_n.

The situation that we are interested in studying is when there is a mismatch between the observations and the underlying stochastic model at small time-scales. It has been shown [27] that there are limitations to the extent Markov models can replicate the data from deterministic systems (or multiscale systems). This limitation, as we shall show, is due to the difference in the curvature of the correlation function close to lag zero.

Therefore, the *empirical covariance estimators* (1.6) based on the approximating process will no longer be optimal under the "regular" conditions, and their study will be pivotal to identifying the good adaptive subsampling schemes.

Chapter 2

Direct Observability of Ornstein-Uhlenbeck Process

2.1 Introduction

We introduce parametric estimation of a stochastic model by considering a specific example. Let $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}_{t\geq 0}, P)$ be a filtered probability space. Consider a time-homogeneous (drift and diffusion coefficients do not depend on the time variable) one-dimensional diffusion process $\{X_t,\ t \in [0,\infty)\}$ whose dynamics are represented by the following stochastic differential equation,

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{2.1}$$

where $\gamma > 0$, $\sigma > 0$ are unknown parameters such that (γ, σ^2) lies in some parameter space $\Theta \subset \mathbb{R}^2$, and W_t is the standard Brownian motion adapted to the filtration $\mathcal{F}_t$.

In this study our objective is to present an approach to estimate the parameters γ, σ^2 under *indirect observability*, i.e., when the underlying process X_t is *unobserved*, and the observed process Y_t^ε is only an approximation of X_t. Before embarking on the case of *indirect observability*, we develop the design of estimation and set up the machinery to perform

asymptotic analysis of the estimators based on discretely sampled *direct* observations.

Direct observability is our benchmark case and our workhorse to illustrate the asymptotic results for parametric estimation. Our objective is to estimate unknown parameters associated to an underlying process based on discrete observations of this process. The estimators should be such that under adequate asymptotics they should lie in an arbitrary small neighborhood of true parameter values. The model thus identified, by the estimates of the unknown parameters, can then be relied upon for understanding the nature of the underlying physical process, and be analyzed for trend, periodicity, prediction, and the like.

The linear SDE defined by (2.1) has a strong solution given by,

$$X_t = X_0 e^{-\gamma t} + \sigma e^{-\gamma t} \int_0^t e^{\gamma s} dW_s, \tag{2.2}$$

with $X_0 \sim \mathbf{N}(0, \sigma^2/2\gamma)$, which is the known invariant distribution for (2.1). The stochastic process $\{X_t : t \geq 0\}$ is called the *Ornstein-Uhlenbeck* (OU) process or the *Gauss-Markov* process. It is stationary, Gaussian, Markovian, and continuous in probability [85, 87] (see appendix 0.11 for definitions). The OU process necessarily, [29, 52, 18], satisfies the linear SDE defined by (2.1).

The properties of the OU process are discussed in the Appendix 0.14. In the next section we describe the estimation procedure for the OU parameters γ and σ^2.

A typical design for estimation is to assume as given a set of discretely sub-sampled observations U_n at equidistant time steps, $\{0 = t_0 < t_1 = \Delta < \ldots < t_N = N\Delta\}$, of length $\Delta > 0$. It may be emphasized that under *direct observability* discrete observations U_n satisfy the SDE (2.1), i.e, $U_n = X_{n\Delta}$. The estimators for γ, σ^2 are functions of the observations U_n, and N, Δ. These estimators are obtained by optimizing some criterion $C_N(\gamma, \sigma^2)$ over the parameter space Θ.

2.2 Likelihood Approach

Here we utilize the maximum likelihood approach to derive estimators for the drift and diffusion coefficients γ, σ^2. The likelihood (or approximate likelihood) function, based on the given observations, provides a criterion that is maximized with respect to the unknown parameters to obtain estimates.

Proposition 2.2.1 (EM discretization) *Consider $(N+1)$ discretely sampled observations $U_n = X_{n\Delta}$ at equidistant time points separated by length $\Delta > 0$. The optimal estimators based on the likelihood function constructed from the Euler-Maruyama (EM) discretization are given by,*

$$\begin{aligned} \hat{\gamma} &= \frac{1}{\Delta}\left[1 - \frac{\hat{r}_1}{\hat{r}_0} + \frac{(\hat{\sigma}^2 - 2\hat{\gamma}U_0^2)}{2\hat{\gamma}N\hat{r}_0}\right], \\ \hat{\sigma}^2 &= \left[\frac{U_N^2 + U_0^2}{(N+1)\Delta} + 2\hat{\gamma}\hat{r}_0\left(\frac{N\,(2-\hat{\gamma}\Delta)}{2(N+1)}\right)\right] / \left[1 + \frac{(1-\hat{\gamma}\Delta)}{\hat{\gamma}(N+1)\Delta}\right], \end{aligned} \tag{2.3}$$

where the standard empirical covariance estimators $\hat{r}_0 = \hat{r}_0(N,\Delta)$, $\hat{r}_1 = \hat{r}_1(N,\Delta)$ are defined by,

$$\begin{aligned} \hat{r}_0(N,\Delta) &= \frac{1}{N}\sum_{n=0}^{N-1} U_n^2, \\ \hat{r}_1(N,\Delta) &= \frac{1}{N}\sum_{n=0}^{N-1} U_{n+1}U_n. \end{aligned} \tag{2.4}$$

Proof. The approximate log likelihood function, as given by (1.5), computed from the EM discretization, in particular, of the SDE (2.1) is defined as,

$$\mathbb{L}_N(\theta) = -\frac{1}{2}\log\left(\frac{\pi\sigma^2}{\gamma}\right) - \frac{\gamma U_0^2}{\sigma^2} - \frac{1}{2}\sum_{n=0}^{N-1}\left[\log\left(2\pi\Delta\sigma^2\right) + \frac{\left(U_{n+1} - U_n + \gamma\Delta U_n\right)^2}{\sigma^2\Delta}\right]. \tag{2.5}$$

The first-order necessary conditions $\nabla \mathbb{L}_N(\theta) = 0$ lead to closed form expressions for the optimal estimators $\hat{\gamma} = \hat{\gamma}(N, \Delta)$ and $\hat{\sigma}^2 = \hat{\sigma}^2(N, \Delta)$ given by (2.3) ■

Alternatively, the transition density of the OU process is known via the exact solution (2.2), therefore, in this case we do not have to use discretization to approximate the likelihood function. From (2.2), the observations U_n sub-sampled from the trajectory of an OU process satisfy the following equation

$$U_{n+1} = e^{-\gamma\Delta} U_n + \sigma e^{-\gamma t_{n+1}} \int_{t_n}^{t_{n+1}} e^{\gamma s} dW_s, \tag{2.6}$$

where the Itô integral $\int_{t_n}^{t_{n+1}} e^{\gamma s} dW_s \sim \mathbf{N}\left(0, (e^{2\gamma t_{n+1}} - e^{2\gamma t_n})/(2\gamma)\right)$, and U_n is observation sampled at time point t_n. In general, for each $t > 0$, the Itô integral $\int_0^t G(s) dW_s$ is a zero-mean Gaussian random variable, for some real-valued deterministic continuous function G on R_+. The variance of the Itô integral is computed using the Itô isometry,

$$E\left[\left(\int_{t_n}^{t_{n+1}} e^{\gamma s} dW_s\right)^2\right] = \int_{t_n}^{t_{n+1}} e^{2\gamma s} ds.$$

Proposition 2.2.2 (Exact likelihood) *Consider* $(N+1)$ *discretely sampled observations* $U_n = X_{n\Delta}$ *at equidistant time points separated by length* $\Delta > 0$. *The process* X_t *is the strictly stationary Ornstein-Uhlenbeck process, and the observations* U_n *satisfy the discrete relation* (2.6). *Then, the optimal estimators* $\hat{\gamma} = \hat{\gamma}(N, \Delta)$, $\hat{\sigma}^2 = \hat{\sigma}^2(N, \Delta)$ *verify the following conditions,*

$$\begin{aligned} \hat{\gamma} &= \Delta^{-1} \log\left(\frac{\hat{r}_0 + N^{-1}\left((\hat{\sigma}^2/2\hat{\gamma}) - U_0^2\right)}{\hat{r}_1}\right), \\ \hat{\sigma}^2 &= \left(\frac{2\hat{\gamma}}{N+1}\right)\left(N\hat{r}_0 + \frac{2\hat{\gamma}\left(U_N^2 e^{2\hat{\gamma}\Delta} - U_0^2\right)}{(e^{2\hat{\gamma}\Delta} - 1)}\right). \end{aligned} \tag{2.7}$$

where $\hat{r}_0 = \hat{r}_0(N, \Delta)$, *and* $\hat{r}_1 = \hat{r}_1(N, \Delta)$ *are the standard empirical covariance estimators*

given by (2.4).

Proof. Based on the discrete equation (2.6) the log likelihood function is given by

$$\mathbb{L}_{N,1}(\theta) = -\frac{1}{2}\log\left(\frac{\pi\sigma^2}{\gamma}\right) - \frac{\gamma U_0^2}{\sigma^2} - \frac{N}{2}\log\left(\frac{\pi\sigma^2}{\gamma}(1-e^{-2\gamma\Delta})\right)\ldots \\ -\frac{\gamma\sum_{n=0}^{N-1}(U_{n+1}-e^{-\gamma\Delta}U_n)^2}{\sigma^2\,(1-e^{-2\gamma\Delta})}. \tag{2.8}$$

The first-order necessary conditions $\nabla\mathbb{L}_{N,1}(\theta) = 0$ lead to closed form expressions for the optimal estimators given by (2.7). ■

It may be noted that here we have assumed the observations U_n are sampled from a strictly stationary OU process. If $X_0 = x_0 \in \mathbb{R}$ then the OU process X_t is only an *asymptotically stationary* Gaussian process. The former case of strict stationarity may be desirable as many results for stationary Gaussian processes [10] are then applicable to our study. However, it may be noted that if we assume that the observations are sampled from asymptotically stationary OU process, then, the likelihood function constructed from the exact solution (2.6) is given as,

$$\mathbb{L}_{N,1}(\theta) = -\frac{N}{2}\log\left(\frac{\pi\sigma^2}{\gamma}(1-e^{-2\gamma\Delta})\right) - \frac{\gamma\sum_{n=0}^{N-1}(U_{n+1}-e^{-\gamma\Delta}U_n)^2}{\sigma^2\,(1-e^{-2\gamma\Delta})}. \tag{2.9}$$

Then, the optimal estimators $\hat{\gamma}$, $\hat{\sigma}^2$, verifying the necessary conditions $\nabla\mathbb{L}_{N,2}(\gamma,\sigma^2) = 0$ are given as,

$$\begin{aligned} \hat{\gamma} &= \frac{1}{\Delta}\log\left(\frac{\hat{r}_0}{\hat{r}_1}\right), \\ \hat{\sigma}^2 &= 2\,\hat{\gamma}\left(\hat{r}_0 + \frac{e^{2\hat{\gamma}\Delta}\left(U_N^2 - U_0^2\right)}{N\,(e^{2\hat{\gamma}\Delta}-1)}\right), \end{aligned} \tag{2.10}$$

where $\hat{r}_0, \hat{r}_1$ are given by (2.4).

The necessary conditions verified by the estimators $\hat{\gamma}$, $\hat{\sigma}^2$ in (2.10) provide explicit expressions for the two estimators, as opposed to the implicit conditions in (2.3) and (2.7). However, the estimators in (2.10), specifically the expression for $\hat{\sigma}^2$, can be further reduced to provide *asymptotically equivalent* estimators, namely,

$$\begin{aligned} \hat{\gamma} &= \frac{1}{\Delta} \log\left(\frac{\hat{r}_0}{\hat{r}_1}\right), \\ \hat{\sigma}^2 &= 2\,\hat{\gamma}\,\hat{r}_0. \end{aligned} \tag{2.11}$$

Comparing $\hat{\sigma}^2$ in (2.10) and (2.11), we obtain that the additional term $\left(\frac{2\hat{\gamma}e^{2\hat{\gamma}\Delta}\left(U_N^2 - U_0^2\right)}{N\left(e^{2\hat{\gamma}\Delta}-1\right)}\right)$ converges to zero in L_2 under appropriate asymptotics.

Lemma 2.2.3 *The L_2 norm of the term $\left(\frac{2\hat{\gamma}e^{2\hat{\gamma}\Delta}\left(U_N^2 - U_0^2\right)}{N\left(e^{2\hat{\gamma}\Delta}-1\right)}\right)$ converges to 0, provided the following conditions hold,*

$$N \to \infty, \qquad \Delta = \Delta(N) \to 0, \qquad N\Delta \to \infty.$$

Proof. We have the following inequality,

$$\left\| \left(\frac{2\hat{\gamma}e^{2\hat{\gamma}\Delta}\left(U_N^2 - U_0^2\right)}{N\left(e^{2\hat{\gamma}\Delta}-1\right)}\right) \right\|_{L_2} = \frac{1}{N\Delta} \left\| \left(\frac{2\hat{\gamma}\Delta e^{2\hat{\gamma}\Delta}\left(U_N^2 - U_0^2\right)}{\left(e^{2\hat{\gamma}\Delta}-1\right)}\right) \right\|_{L_2} \leq \frac{cte}{N\Delta},$$

where *cte* is an arbitrary positive constant. The uniform bound is derived from the Gaussianity of the subsampled process U_n, and using similar arguments as presented in lemma 2.6.3. ■

Similarly, we can show that the estimators in (2.3) and (2.7) are also asymptotically equivalent to (2.11). The final estimators obtained in (2.11), and the covariance function

for the stationary OU process, given as,

$$K(h) = E[X_t X_{t+h}] = \frac{\sigma^2}{2\gamma} e^{-\gamma h}, \qquad \forall h \geq 0, \tag{2.12}$$

inspire the following approach to arrive at the estimators (2.11). This approach will be extended to more general framework (Section 6), and exploited under the more complex scenario of *indirect observability*. An alternative approach to arrive at the estimators (2.11) is presented in the next section utilizing the second order moment conditions.

2.3 Estimating the Covariance Function

Consider the covariance function (2.12) of the OU process for $h = 0$, and $h = \Delta > 0$, namely,

$$r_0 = K(0) = \frac{\sigma^2}{2\gamma}, \qquad r_1 = K(\Delta) = r_0 e^{-\gamma\Delta}. \tag{2.13}$$

The relation (2.13) between OU parameters γ, σ^2 and the covariances $r_0 = r_0(\Delta)$ and $r_1 = r_1(\Delta)$ implies that

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

where the smooth functions g, s are given by

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

Consider $(N+1)$ observations $U_n = X_{n\Delta}$, $n = 0, \ldots, N$; subsampled directly from the OU process with time step length $\Delta > 0$. The estimators $\hat{\gamma} = \hat{\gamma}(N, \Delta)$ and $\hat{\sigma}^2 = \hat{\sigma}^2(N, \Delta)$ are

given by

$$\hat{\gamma} = g(\hat{r}_0, \hat{r}_1) = \frac{1}{\Delta} \log\left(\frac{\hat{r}_0}{\hat{r}_1}\right),$$
$$\hat{\sigma}^2 = s(\hat{r}_0, \hat{r}_1) = \frac{2\hat{r}_0}{\Delta} \log\left(\frac{\hat{r}_0}{\hat{r}_1}\right), \tag{2.14}$$

where $\hat{r}_0$, $\hat{r}_1$ are the standard empirical covariance estimators (2.4), which yields the explicit estimators given in (2.11). Since, two parameters γ, σ^2 characterize the OU process, we require only two moment conditions (2.13) to arrive at the expressions for the estimators.

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

We showed (lemma 2.2.3) that the estimators for γ and σ^2 in (2.11) are *asymptotically equivalent*, in particular, to the estimators given by (2.7). Therefore, for our study we fix the estimators given by (2.14) (or (2.11)) for the drift and diffusion coefficients γ, σ^2 in the OU process. The estimators $\hat{\gamma}$, $\hat{\sigma}^2$ in (2.14) are expressed as differentiable functions of the standard empirical covariance estimators $\hat{r}_0, \hat{r}_1$. This representations is pivotal in our study of the asymptotic properties of the estimators. This is mainly because some asymptotic properties, as shown below, are preserved by the smooth functions. It is known [10], for instance, that the consistency and asymptotic Gaussianity, under *direct observability*, of the standard empirical covariance estimators $\hat{r}_k(N, \Delta)$ is preserved under smooth functions.

In chapter 3 we show that asymptotic consistency holds under *indirect observability* of the OU process. But, first we present the results on consistency for the benchmark case of direct observability.

2.4 Asymptotic Conditions on N and Δ

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

The asymptotic properties of estimators $\hat{\gamma}$, $\hat{\sigma}^2$ when the observations U_n are sampled from the true OU process are now presented under the two types of subsampling schemes.

Definition 2.4.1 *We say that we have a* ***Fixed-rate Subsampling*** *scheme when the time-step $\Delta > 0$ between observations remains fixed as $N \to \infty$.*

Definition 2.4.2 *We say that we have an* ***Adaptive Subsampling*** *scheme when the time-step between observations is a function $\Delta(N)$ of N tending to 0 as $N \to \infty$, and we then always impose the condition $N\Delta(N) \to \infty$.*

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

Since, OU process is Gaussian, and without loss of generality it is assumed that we are subsampling from the strictly stationary version of the OU process, therefore, we show that the estimators converge to the true values in L^2 under both the *subsampling schemes*. We require the result in proposition 2.4.3 to compute the L^2 norm of the *empirical covariance estimators* $\hat{r}_k(N, \Delta)$, as defined in (2.20), which will be crucial to further compute the L^2 norm of the OU estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ given by (2.14).

These results are easily extended to the situation where the OU process is only asymptotically stationary because convergence to its stationary distribution is exponentially fast. Hence, the OU process observed for $t \geq t_0$ such that $t_0 \gg (\gamma^{-1})$, may essentially be considered as stationary.

Proposition 2.4.3 (Finite N) *Let $(U_n)_{n\in\mathbb{Z}}$ be a centered stationary Gaussian process. Define the covariances r_k for each $k \in \mathbb{Z}$ by*

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

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

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

By stationarity of the process U_n the empirical covariance estimators are unbiased, i.e., $E[\hat{r}_k(N)] = r_k$, for each $k \in \mathbb{Z}$. Then, for each pair of non-negative integers k, q, the covariance of the estimators is given by

$$Cov(\hat{r}_k(N), \hat{r}_q(N)) = (1/N)\sum_{j=-(N-1)}^{N-1} f(j) - (1/N^2)\sum_{j=1}^{N-1} j(f(j) + f(-j)), \tag{2.15}$$

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

Proof. The covariance

$$Cov(\hat{r}_k(N), \hat{r}_q(N)) = E\left[\hat{r}_k(N)\hat{r}_q(N)\right] - r_k r_q,$$

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

$$N^2 E\left[\hat{r}_k(N)\hat{r}_q(N)\right] = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} E\left[U_m U_{m+k} U_n U_{n+q}\right]. \tag{2.16}$$

A well known result for the Gaussian random variables gives us the 4th-order moments in

terms of the 2nd-order moments, precisely

$$E\left[U_m U_{m+k} U_n U_{n+q}\right] = r_k r_q + f(m-n), \tag{2.17}$$

where $f(m-n) = \left(r_{m-n} r_{m-n+k-q} + r_{m-n-q} r_{m-n+k}\right)$. Substituting (2.17) in (2.16) we obtain,

$$N^2 E\left[\hat{r}_k(N)\hat{r}_q(N)\right] = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left[r_k r_q + f(m-n)\right].$$

Therefore, the covariance is given by,

$$Cov(\hat{r}_k(N), \hat{r}_q(N)) = N^{-2}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m-n).$$

The double sum can be simplified to give,

$$Cov(\hat{r}_k(N), \hat{r}_q(N)) = N^{-1}\sum_{j=-(N-1)}^{N-1} f(j) - N^{-2}\sum_{j=1}^{N-1} j(f(j)+f(-j)),$$

where $f(j) = (r_j r_{j+k-q} + r_{j+k} r_{j-q})$, which gives the required expression. ■

In the next two sections we will study the asymptotic properties of the estimators (2.14) under fixed-rate, and adaptive subsampling schemes, and the result from proposition 2.4.3 will be crucial.

2.5 Fixed-rate Subsampling

Assume *fixed-rate subsampling* for the observations $U_n = X_{n\Delta}$ subsampled from the OU process, so that $\Delta > 0$ is fixed. The stationary covariances $r_k = r_k(\Delta)$ are, thus, given by

$$r_k = E[U_{n+k}U_n] = \frac{\sigma^2}{2\gamma} e^{-\gamma k \Delta}, \tag{2.18}$$

which implies the relation

$$\sum_{k\in\mathbb{Z}} |k||r_k(\Delta)| < \infty. \tag{2.19}$$

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

$$\hat{r}_0 = \hat{r}_0(N,\Delta) = \frac{1}{N}\sum_{n=0}^{N-1} U_n^2, \qquad \hat{r}_1 = \hat{r}_1(N,\Delta) = \frac{1}{N}\sum_{n=0}^{N-1} U_{n+1}U_n. \tag{2.20}$$

Since the covariances r_k verify (2.19), known results [10] on discrete-time stationary Gaussian processes show that for each fixed $\Delta > 0$ as $N \to \infty$, the covariance estimators $\hat{r}_k(N,\Delta)$ are the best estimators of the r_k, they are *consistent* (i.e., converges almost surely to the true r_k), and *asymptotically efficient* (i.e., the asymptotic variances of $\hat{r}_k(N,\Delta)$ attain the Cramér-Rao bound). We also know (see [10, Chapter X]) that as $N \to \infty$, the random vectors

$$\sqrt{N}\,[\hat{r}_0(N,\Delta) - r_0, \hat{r}_1(N,\Delta) - r_1]$$

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

$$\Gamma_{st} = \sum_{m\in\mathbb{Z}} \left(r_m r_{m-s+t} + r_{m-s}r_{m+t}\right), \tag{2.21}$$

with r_m given by (2.18), and hence, the covariance matrix Γ is given by

$$\begin{aligned}
\Gamma_{00} &= 2r_0^2(1+e^{-2\gamma\Delta})/\left(1-e^{-2\gamma\Delta}\right), \\
\Gamma_{11} &= r_0^2\left(1+4e^{-2\gamma\Delta}-e^{-4\gamma\Delta}\right)/\left(1-e^{-2\gamma\Delta}\right), \\
\Gamma_{01} &= \Gamma_{10} = 4r_0^2 e^{-\gamma\Delta}/\left(1-e^{-2\gamma\Delta}\right),
\end{aligned} \tag{2.22}$$

where $r_0 = (\sigma^2/2\gamma)$.

Alternately, the result for finite N from proposition (2.4.3) can be used to find the L_2

norms of the empirical covariance estimators $\hat{r}_0$, $\hat{r}_1$. Then, we may be able to infer the L_2 convergence of the estimators $\hat{\gamma}$, $\hat{\sigma}^2$.

Proposition 2.5.1 (L^2-norm for finite N) *Consider $(N+1)$ discrete observations $U_n = X_{n\Delta}$ subsampled from the trajectory of an OU process at fixed time step length $\Delta > 0$. Define the standard empirical covariance estimators $(\hat{r}_k)_{k=0,1}$ by,*

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

Then, for each fixed N, $\Delta > 0$, the L_2-norms $J_0 = \|\hat{r}_0(N,\Delta) - r_0\|_{L_2}^2$, and $J_1 = \|\hat{r}_1(N,\Delta) - r_1\|_{L_2}^2$ are given by

$$\begin{aligned} J_0 &= J_0(N,\Delta) = \frac{2r_0^2(1+e^{2\gamma\Delta})}{N(e^{2\gamma\Delta}-1)} + \frac{4r_0^2 e^{2\gamma\Delta}\left(e^{-2\gamma N\Delta}-1\right)}{N^2(e^{2\gamma\Delta}-1)^2}, \\ J_1 &= J_1(N,\Delta) = \frac{r_0^2(e^{4\gamma\Delta}+4e^{2\gamma\Delta}-1)}{Ne^{2\gamma\Delta}(e^{2\gamma\Delta}-1)} + \frac{4r_0^2 e^{2\gamma\Delta}\left(e^{-2\gamma N\Delta}-1\right)}{N^2(e^{2\gamma\Delta}-1)^2}. \end{aligned} \tag{2.23}$$

Proof. Let $J_k(N,\Delta) = E\left[\left(\hat{r}_k(N,\Delta) - r_k(\Delta)\right)^2\right]$, for each non-negative integer k. The empirical covariance estimator $\hat{r}_k$, and covariance $r_k(\Delta)$ are defined as

$$\hat{r}_k = \hat{r}_k(N,\Delta) = \frac{1}{N}\sum_{n=0}^{N-1} U_n U_{n+k}, \qquad r_k(\Delta) = (\sigma^2/2\gamma)e^{-\gamma k\Delta}.$$

Then, using the expression (2.15) in proposition 2.4.3, such that $J_k = C_{kk}$, and $f(j) = \left(r_j^2 + r_{j-k}r_{j+k}\right)$, we obtain,

$$\begin{aligned} J_k(N,\Delta) &= \frac{r_0^2\left[1+e^{-2\gamma\Delta}+(2k+1)e^{-2\gamma k\Delta}-(2k-1)e^{-2\gamma(k+1)\Delta}\right]}{N(1-e^{-2\gamma\Delta})}\ldots \\ &-\frac{2r_0^2}{N^2}\left[\frac{e^{-2\gamma\Delta}(1+e^{-2\gamma k\Delta}-2e^{-2\gamma N\Delta})}{(1-e^{-2\gamma\Delta})^2} + \frac{ke^{-2\gamma k\Delta}}{(1-e^{-2\gamma\Delta})} + \frac{k(k-1)e^{-2\gamma k\Delta}}{2}\right]. \end{aligned} \tag{2.24}$$

Substituting $k = 0, 1$ in (2.24) gives us the required expressions. ■

From the L_2 norms J_0, J_1 (2.23), we obtain that, as $N \to \infty$, the empirical covariance estimators $\hat{r}_0(N, \Delta)$, $\hat{r}_1(N, \Delta)$ converge in L_2, respectively, to $r_0(\Delta)$ and $r_1(\Delta)$. We now study the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ given by,

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

which have the explicit expressions

$$\hat{\gamma} = -\frac{1}{\Delta} \log \left(\frac{\hat{r}_1(N, \Delta)}{\hat{r}_0(N, \Delta)} \right), \qquad \hat{\sigma}^2 = 2\hat{\gamma}\, \hat{r}_0(N, \Delta). \tag{2.25}$$

Proposition 2.5.2 (Fixed-rate asymptotics for $\hat{\gamma}$ and $\hat{\sigma}^2$) *Consider an OU-process X_t directly observed at times $t = n\Delta$, $n = 0, \ldots, N$, subsampling at fixed rate $\Delta > 0$. Then as $N \to \infty$, the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ of γ and σ^2 are consistent (almost surely). Moreover $\sqrt{N}(\hat{\gamma} - \gamma)$ and $\sqrt{N}(\hat{\sigma}^2 - \sigma^2)$ are asymptotically Gaussian with limit variances v_γ and v_{σ^2} given by*

$$\begin{aligned} v_\gamma &= \left(\frac{e^{-2\gamma\Delta} + e^{2\gamma\Delta} - 2}{\Delta^2 (1 - e^{-2\gamma\Delta})} \right), \\ v_{\sigma^2} &= 4r_0^2 \left(\frac{2(1 + \gamma\Delta)^2 (1 + e^{-2\gamma\Delta}) - 8\gamma\Delta + e^{2\gamma\Delta} - e^{-2\gamma\Delta} - 4}{\Delta^2 (1 - e^{-2\gamma\Delta})} \right). \end{aligned} \tag{2.26}$$

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

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

Since F is twice continuously differentiable in the neighborhood of (r_0, r_1) for each $\Delta > 0$, it follows from [10, Chapter X] that the estimator $\hat{\nu} = F(\hat{r}_0(N, \Delta), \hat{r}_1(N, \Delta))$ is a consistent estimator of $\nu = F(r_0, r_1)$. Also, the distribution of $\sqrt{N}\,(\hat{\nu} - \nu)$ converges, as $N \to \infty$,

towards the 2-dimensional centered Gaussian distribution with covariance matrix

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

where, for each fixed $\Delta > 0$, the (2×2) matrix $A = \nabla F(r_0, r_1)$ is the differential of F at true covariances (r_0, r_1), A^T denotes the transpose of A, and the covariance matrix Γ is given by (2.22). This says exactly that $\hat{\gamma}$ and $\hat{\sigma}^2$ are *consistent and asymptotically Gaussian estimators* of γ and σ^2, and that Σ_F is the limit covariance matrix of the random vector

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

■

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

In many practical situations the observations are sampled on a fixed global time span $T = N\Delta$, therefore, taking $N \to \infty$ imposes $\Delta \to 0$. The small time step length Δ leads to blowing up of the asymptotic variance of, for instance, $\sqrt{N}(\hat{\gamma} - \gamma)$ given by (2.26), namely,

$$v_\gamma = \left(\frac{e^{-2\gamma\Delta} + e^{2\gamma\Delta} - 2}{\Delta^2(1 - e^{-2\gamma\Delta})}\right) = \frac{2\gamma}{\Delta}\left(1 + O(\Delta)\right),$$

and the L_2-errors do not converge to 0 for $T = N\Delta$ bounded. Therefore, the correct scaling to study the asymptotic properties of the estimators will be $T = N\Delta$, and so adaptive subsampling schemes (see definition 2.4.2) are a natural progression from here.

2.6 Adaptive Subsampling

We now study the consistency of the estimators $\hat{\gamma} = \hat{\gamma}(N, \Delta)$ and $\hat{\sigma}^2 = \hat{\sigma}^2(N, \Delta)$ under *adaptive subsampling scheme* (see definition 2.4.2).

Proposition 2.6.1 (Asymptotics of the Covariances) *Consider an adaptive subsampling scheme where we have N observations $U_n = X_{n\Delta}$ of the stationary OU process X_t at time intervals of length $\Delta = \Delta(N)$ depending on N. We assume (see definition 2.4.2)*

$$\Delta \to 0, \quad N\Delta \to \infty. \tag{2.27}$$

The true covariances $r_k = r_k(\Delta)$ of the process U_n are now functions of N still given by (2.18)*. Hence as $N \to \infty$, and for each $k \geq 0$, $r_k(\Delta(N)) \to (\sigma^2/2\gamma)$.*

Then, under condition (2.27)*, and for each $k \geq 0$, the empirical covariances $\hat{r}_k(N, \Delta)$ converge in L_2 to $(\sigma^2/2\gamma)$. Moreover, for each $k \geq 0$ the L^2-norms of the variables $\sqrt{N\Delta}\,(\hat{r}_k(N, \Delta) - r_k)$ converge to $\left(\sigma^2/\gamma\sqrt{2\gamma}\right)$.*

Proof. The associated speed of convergence to zero in L_2 for the difference $(\hat{r}_k(N, \Delta) - r_k)$ under adaptive subsampling scheme can be computed directly, as outlined here for $k = 0$ and $k = 1$. Let $J_k = E\left[(\hat{r}_k - r_k)^2\right]$, then, using the expression for J_k given by (2.24) in proposition 2.5.1 we obtain for $k = 0, 1$,

$$J_0 = J_0(N, \Delta) = \frac{2r_0^2(1 + e^{2\gamma\Delta})}{N(e^{2\gamma\Delta} - 1)} + \frac{4r_0^2 e^{2\gamma\Delta}\left(e^{-2\gamma N\Delta} - 1\right)}{N^2(e^{2\gamma\Delta} - 1)^2},$$

$$J_1 = J_1(N, \Delta) = \frac{r_0^2(e^{4\gamma\Delta} + 4e^{2\gamma\Delta} - 1)}{Ne^{2\gamma\Delta}(e^{2\gamma\Delta} - 1)} + \frac{4r_0^2 e^{2\gamma\Delta}\left(e^{-2\gamma N\Delta} - 1\right)}{N^2(e^{2\gamma\Delta} - 1)^2}.$$

Under the conditions (2.27), since, $\gamma > 0$, the expressions $\left(\Delta/(e^{2\gamma\Delta}-1)\right) \to (2\gamma)^{-1}$, and $N(e^{2\gamma\Delta}-1) \to \infty$, which proves the following convergence,

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

This concludes the proof. ■

Proposition 2.6.2 *For each N, Δ, the random variables $Z_0 = Z_0(N,\Delta)$ and $Z_1 = Z_1(N,\Delta)$ defined by,*

$$Z_0 = \frac{(\hat{r}_0 - r_0)}{\sqrt{J_0}}, \qquad Z_1 = \frac{(\hat{r}_1 - r_1)}{\sqrt{J_1}}, \tag{2.28}$$

have mean 0, variance 1, and covariance $E[Z_0 Z_1] = J_{01}/(\sqrt{J_0 J_1})$, where J_0, J_1, J_{01} are given by (2.23), (2.30). *Then, under conditions* (2.27) *the following first-order L_2 approximations for the empirical covariances $\hat{r}_k$ hold,*

$$\begin{aligned} \hat{r}_0 &= r_0 + \frac{r_0}{\sqrt{N\Delta}}\sqrt{\frac{2}{\gamma}}Z_0 + \frac{Z_0}{\sqrt{N\Delta}} \times O\left(\Delta^2 + \frac{1}{N\Delta}\right), \\ \hat{r}_1 &= r_1 + \frac{r_0}{\sqrt{N\Delta}}\sqrt{\frac{2}{\gamma}}Z_1 + \frac{Z_1}{\sqrt{N\Delta}} \times O\left(\Delta^2 + \frac{1}{N\Delta}\right), \end{aligned} \tag{2.29}$$

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

Proof. The exact expression for $J_{01} = J_{01}(N,\Delta)$ derived using proposition 2.4.3, is given by

$$J_{01} = \frac{4r_0^2 e^{\gamma\Delta}}{N(e^{2\gamma\Delta}-1)} - \frac{2r_0^2 e^{\gamma\Delta}(e^{2\gamma\Delta}+1)(1-e^{-2\gamma N\Delta})}{N^2(e^{2\gamma\Delta}-1)^2}. \tag{2.30}$$

Applying Taylor expansions to J_k as given by (2.23) and J_{01} given by (2.30), we obtain

$$\begin{aligned} J_0 &= \frac{2r_0^2}{\gamma N\Delta}\left(1+\frac{\gamma^2\Delta^2}{3}-\frac{1+O(\Delta^2)}{2\gamma N\Delta}+O(\Delta^4)\right), \\ J_1 &= \frac{2r_0^2}{\gamma N\Delta}\left(1-\frac{2\gamma^2\Delta^2}{3}+\gamma^3\Delta^3-\frac{1+O(\Delta^2)}{2\gamma N\Delta}+O(\Delta^4)\right), \\ J_{01} &= \frac{2r_0^2}{\gamma N\Delta}\left(1-\frac{\gamma^2\Delta^2}{6}-\frac{1+O(\Delta^2)}{2\gamma N\Delta}+O(\Delta^4)\right). \end{aligned} \tag{2.31}$$

Substituting in (2.28) the above expressions for J_0, J_1 gives the required L_2-approximations as expressed in (2.29). ■

Define the random variable Z_k, for any integer $k \geq 0$, as $Z_k = (\hat{r}_k - r_k)/\sqrt{J_k}$, where $J_k = C_{kk}$ is given by (2.24). The next lemma will be needed to prove the consistency of $\hat{\gamma}$ and $\hat{\sigma}^2$.

Lemma 2.6.3 *For each integer $k \geq 0$, consider a random variable $V_k = V_k(\theta)$ given by,*

$$V_k = \left(\frac{a_k Z_k}{1+a_k\theta Z_k}\right)^2, \tag{2.32}$$

where $Z_k = (\hat{r}_k - r_k)/\sqrt{J_k}$, $\theta \in (0,1)$ and $a_k = e^{\gamma k\Delta}\sqrt{J_k}/r_0$, such that $J_k \sim O(1/N\Delta)$. Then, under the condition (2.27), *$\|V_k\|_{L_2} \to 0$ with a speed proportional to $1/N\Delta$.*

Proof. The L^2-norm is given by $\|V_k\|_{L_2}^2 = E\left[(a_k Z_k/(1+a_k\theta Z_k))^4\right]$. Since, the tails of the density for Z_k decay exponentially fast, we have for any $M \gg 1$, $P\left\{(1+a_k\theta Z_k)^{-1} > M\right\} < e^{-(C_1\sqrt{N\Delta}/\theta)}$, where C_1 is a positive constant. Also,

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

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

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

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

Proposition 2.6.4 (Consistency of $\hat{\gamma}$ and $\hat{\sigma}^2$) *Consider the adaptive sub-sampling scheme providing N observations $U_n = X_{n\Delta}$ of the stationary OU process X_t at time intervals of length $\Delta = \Delta(N)$. Define the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ by formula* (2.25).

Then, under the conditions

$$\Delta \to 0, \quad N\Delta \to \infty, \tag{2.33}$$

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

Moreover, given (2.33), *the L_2 norms of the variables $\sqrt{N\Delta}\,(\hat{\gamma} - \gamma)$, and $\sqrt{N\Delta}\,\left(\hat{\sigma}^2 - \sigma^2\right)$ converge, respectively, to $\sqrt{2\gamma}$ and 0. Therefore, the estimators converge to the true values with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$. In particular, under stronger conditions,*

$$\Delta \to 0, \quad N\Delta^2 \to \infty, \tag{2.34}$$

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

Proof. From (2.28) we obtain, $\hat{r}_0 = r_0 + \sqrt{J_0} Z_0$, and $\hat{r}_1 = r_1(\Delta) + \sqrt{J_1} Z_1$, which we substitute in

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

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

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

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

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

Using lemma 2.6.3, under the conditions (2.33), and using Taylor expansion we obtain that the following holds in L_2,

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

where the random remainder terms V_0, V_1 are given by (2.32) such that the L_2 norms $\|V_1 - V_0\|_{L_2} \sim O(1/N)$, $\|V_0\|_{L_2} \sim O(1/N\Delta)$, and $\|V_1\|_{L_2} \sim O(1/N\Delta)$.

Let the random variable Z_γ be defined as

$$Z_\gamma = \left(e^{\gamma\Delta}\sqrt{J_1}Z_1 - \sqrt{J_0}Z_0\right)/(\Delta r_0).$$

The L_2 norm of Z_γ is given by,

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

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

$$\hat{\gamma} = \gamma - Z_\gamma + R_\gamma \times O\left(\frac{1}{N\Delta}\right), \tag{2.35}$$

where the random remainder term $R_\gamma = R_\gamma(N, \Delta)$ is uniformly bounded in L_2 norm. Therefore, under the conditions (2.33), the estimator $\hat{\gamma} \to \gamma$ in L_2 with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$ such that,

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

The diffusion estimator $\hat{\sigma}^2 = 2\hat{\gamma}\hat{r}_0$ by (2.25). Let the random variable Z_{σ^2} be defined as follows,

$$Z_{\sigma^2} = (2/\Delta)\left(e^{\gamma\Delta}\sqrt{J_1}Z_1 - (1+\gamma\Delta)\sqrt{J_0}Z_0\right),$$

then its L_2 norm is given by,

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

Hence, using (2.28) and (2.35) we obtain,

$$\hat{\sigma}^2 = \sigma^2 - Z_{\sigma^2} + R_{\sigma^2} \times O\left(\frac{1}{N\Delta}\right), \tag{2.36}$$

where the random remainder term $R_{\sigma^2} = R_{\sigma^2}(N, \Delta)$ is uniformly bounded in L_2 norm. Therefore, under the conditions (2.33), $\hat{\sigma}^2 \to \sigma^2$ in L_2. Moreover, under the conditions (2.34), the following convergence holds,

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

■

To summarize, when the observations are directly extracted from a stationary OU process then, under the *fixed rate sub-sampling scheme* the approximate MLEs (2.25) for the parameters of the OU process are consistent and asymptotically Gaussian. The L_2-speed

of convergence for the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ as $N \to \infty$ is proportional to $1/\sqrt{N}$ for each fixed $\Delta > 0$.

Under the *adaptive sub-sampling scheme* (2.33), the estimators $\hat{\gamma}$ and $\hat{\sigma}^2$ are asymptotically consistent estimators of γ, σ^2. The usual L_2-speed of convergence to true values, proportional to $1/\sqrt{N\Delta}$, is achievable for the estimators $\hat{\gamma}$, $\hat{\sigma}^2$. In fact for the diffusion estimator $\hat{\sigma}^2$, under stronger conditions on N, Δ, one can achieve a faster L_2-speed of convergence proportional to $1/\sqrt{N}$.

In the next chapter we introduce a more common and more complex scenario in which only indirect observations of the underlying OU process X_t are available, and are generated by another process Y_t^ε which is not identical to X_t, but is simply close to X_t in L_2. In this case *subsampling will become an essential tool to generate consistent estimators* of the underlying parameters. The objective will be to characterize the asymptotic conditions on the number of observations $N = N(\varepsilon) \to \infty$, and the time step length $\Delta = \Delta(\varepsilon) \to 0$, such that $N\Delta \to \infty$, as the small parameter $\varepsilon \to 0$.

Chapter 3

Indirect Observability of Ornstein-Uhlenbeck Process

3.1 Introduction

Estimation of parameters in an underlying stochastic process by observing only an approximating process is the main focus of the thesis. In this chapter we illustrate the *indirect observability framework* for parametric estimation of Ornstein-Uhlenbeck process. The study provides explicit subsampling criteria which guarantee asymptotic consistency of the estimators based on the approximate process.

The *indirect observability* framework refers to the parametric estimation of an underlying *unobservable* stochastic process $X_t = X_t(\theta)$ by observing a process Y_t^ε, which approximates X_t. Consider the Ornstein-Uhlenbeck process X_t identified by the following stochastic differential equation,

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{3.1}$$

where $\gamma, \sigma > 0$ are unknown parameters, such that (γ, σ^2) lies in some parameter space $\Theta \subset \mathbb{R}$. Assume that the process X_t is unobservable. The observable approximating process

Y_t^ε is obtained by averaging the process X_t given by (3.1) over a small moving window of length $\varepsilon > 0$, namely,

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

This chapter introduces the above pair (X_t, Y_t^ε) in detail, and presents the asymptotic results thus obtained for the estimators under this specific example of indirect observability framework. As seen below, this example provides us with a rigorous formulation of the subsampling issue in terms of exact mathematical relations.

First candidate for the approximating process Y_t^ε is obtained by averaging the OU process over a small moving window of length $\varepsilon > 0$. The averaging parameter ε is not only crucial in describing the approximating process, but also for the complete characterization of the subsampling schemes for consistent estimation of parameters γ and σ^2. The process Y_t^ε computed by averaging the OU process X_t is called the *Smoothed Ornstein-Uhlenbeck* process, denoted by SOU process.

The subsampling schemes presented in this chapter are extended to more general stochastic models in chapter 6. More complex examples are considered in chapter 4 where the results obtained in this chapter, in particular, by studying the impact of SOU process on the estimators, are used to infer the optimal subsampling schemes.

3.2 Smoothed Ornstein-Uhlenbeck Process

Definition 3.2.1 *The process Y_t^ϵ called the Smoothed Ornstein-Uhlenbeck process, also denoted as SOU process, is obtained by averaging the process X_t over a sliding window of fixed length $\epsilon > 0$, so that*

$$Y_t^\epsilon = \frac{1}{\epsilon} \int_{t-\epsilon}^{t} X_s ds, \tag{3.2}$$

where X_t is the centered stationary Gaussian OU process given by (3.1)*.*

Y_t^ϵ is a centered stationary Gaussian process with a.s. differentiable trajectories. Due

to locally integrating the process X_t, the process Y_t^ε is no longer Markovian. Since, the conditional probability $P[Y_t^\varepsilon \in A/\mathcal{F}_s] \neq P[Y_t^\varepsilon \in A/X_s]$ for all $s \in (t-\varepsilon, t)$.

Proposition 3.2.2 (Stationary Moments) *The covariance function of* Y_t^ϵ (3.2) *at time lag* h *is given by,*

$$K^\epsilon(h) = E[Y_t^\epsilon Y_{t+h}^\epsilon] = \frac{1}{\epsilon}\left(\int_{t+h-\epsilon}^{t+h} E[X_s Y_t^\epsilon]ds\right). \tag{3.3}$$

Then, the explicit expressions for $K^\epsilon(h)$*, for* $h \geq 0$*, are given by,*

$$K^\epsilon(h) = \begin{cases} \dfrac{\sigma^2}{2\gamma^3\epsilon^2} e^{-\gamma h}\left(e^{-\gamma\epsilon} + e^{\gamma\epsilon} - 2\right). & h \geq \epsilon, \\ \\ \dfrac{\sigma^2}{2\gamma^3\epsilon^2} e^{-\gamma h}\left(2\gamma(\epsilon - h)e^{\gamma h} + e^{-\gamma\epsilon}(e^{2\gamma h} + 1) - 2\right). & h < \epsilon. \end{cases} \tag{3.4}$$

In particular, we have

$$K^\epsilon(0) = \frac{\sigma^2}{\gamma^3\epsilon^2}(e^{-\gamma\epsilon} - 1 + \gamma\epsilon). \tag{3.5}$$

Proof. As is well known, we may in this Gaussian context freely commute expectation signs and integral signs, so that the computation of $K^\epsilon(h)$ boils down to computing simple deterministic integrals of the explicit stationary covariance function of X_t. Since, $E[X_s] = 0$; therefore, the mean of Y_t^ϵ is given by

$$E\left[Y_t^\epsilon\right] = \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} E[X_s]ds = 0.$$

The covariance function of Y_t^ϵ at time lag h is defined to be

$$K^\epsilon(h) = E[Y_{t+h}^\varepsilon Y_t^\epsilon] = \frac{1}{\epsilon}\left(\int_{t+h-\epsilon}^{t+h} E[X_s Y_t^\epsilon]ds\right). \tag{3.6}$$

The stationary covariance of X_t is given by

$$E[X_s X_u] = (\sigma^2/2\gamma)e^{-\gamma(s-u)},$$

therefore, for each $s \in [t+h-\epsilon, t+h]$ such that $h \geq \epsilon$, we get

$$E[X_s Y_t^\epsilon] = \frac{\sigma^2}{2\gamma}\frac{e^{-\gamma s}}{\gamma\epsilon}\left(e^{\gamma t} - e^{\gamma(t-\epsilon)}\right). \tag{3.7}$$

Using (3.6) and (3.7) the covariance function $K^\epsilon(h)$, for $h \geq \epsilon$ is given by,

$$K^\epsilon(h) = \frac{\sigma^2}{2\gamma}e^{-\gamma h}\left(\frac{e^{-\gamma\epsilon} + e^{\gamma\epsilon} - 2}{\gamma^2\epsilon^2}\right).$$

For the case when $0 \leq h < \epsilon$ we re-write (3.6) as

$$K^\epsilon(h) = \frac{1}{\epsilon}\left(\int_{t+h-\epsilon}^{t} E[X_s Y_t^\epsilon]ds + \int_t^{t+h} E[X_s Y_t^\epsilon]ds\right). \tag{3.8}$$

Using the definition of Y_t^ε and the stationary covariances of X_t, for each $s \in [t+h-\epsilon, t]$ such that $0 \leq h < \epsilon$, we get

$$E[X_s Y_t^\epsilon] = \frac{\sigma^2}{2\gamma}\frac{1}{\gamma\epsilon}\left(2 - e^{\gamma(t-\epsilon)}e^{-\gamma s} - e^{-\gamma t}e^{-\gamma s}\right). \tag{3.9}$$

Using (3.7), (3.8) and (3.9) we get the required covariance when $0 \leq h < \epsilon$, namely

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

■

Therefore, the correlation function CF_{sou} of Y^{ϵ} is defined as,

$$CF_{sou}(h) = \begin{cases} \dfrac{1}{2}e^{-\gamma h}\left(\dfrac{e^{-\gamma\epsilon}+e^{\gamma\epsilon}-2}{e^{-\gamma\epsilon}-1+\gamma\epsilon}\right), & h \geq \epsilon, \\ \\ \dfrac{1}{2}e^{-\gamma h}\left(\dfrac{2\gamma(\epsilon-h)e^{\gamma h}+e^{-\gamma\epsilon}(e^{2\gamma h}+1)-2}{e^{-\gamma\epsilon}-1+\gamma\epsilon}\right), & 0 \leq h < \epsilon. \end{cases} \tag{3.10}$$

Now, we analyze the correlation function (3.10) that reveals the fundamental differences between the SOU process and the OU process which make estimation under *indirect observability* a difficult problem.

The correlation function obtained for the SOU process has a typical feature, which arises in many other complex models, for instance, the multiscale dynamics introduced in chapter 4, and dynamical systems (see [27] and references therein). This typical feature is the differentiability of the correlation function, in particular, at time lag $h = 0$. The derivative of the correlation function is given by,

$$\frac{d\left(CF_{sou}(h)\right)}{dh} = \begin{cases} -\dfrac{\gamma}{2}e^{-\gamma h}\left(\dfrac{e^{-\gamma\epsilon}+e^{\gamma\epsilon}-2}{e^{-\gamma\epsilon}-1+\gamma\epsilon}\right), & h \geq \epsilon, \\ \\ \dfrac{\gamma}{2}\left(\dfrac{2e^{-\gamma h}+e^{-\gamma\epsilon}(e^{\gamma h}-e^{-\gamma h})-2}{e^{-\gamma\epsilon}-1+\gamma\epsilon}\right), & 0 \leq h < \epsilon. \end{cases} \tag{3.11}$$

The derivative at lag zero is equal to zero, this is in contrast with the correlation function of the OU process $CF_{ou}(h) = e^{-\gamma|h|}$, which is non-differentiable at lag zero.

Definition 3.2.3 *A stochastic process Z_t is differentiable in the mean square sense, with derivative given by a unique process $\dot{Z}_t$, if $E\left[\left((Z_{t+h}-Z_t)/h-\dot{Z}_t\right)^2\right]$ vanishes in the limit as $h \to 0$. The derivative in the classical sense is a special case of this definition.*

Result 3.2.4 *It can be shown [88] that a stationary random process is differentiable in the mean square sense if and only if its associated correlation function is twice differentiable at*

lag zero.

The second derivative of the correlation function (3.10) is given by,

$$\frac{d^2\left(CF_{sou}(h)\right)}{dh^2} = \begin{cases} \dfrac{\gamma^2}{2} e^{-\gamma h}\left(\dfrac{e^{-\gamma\epsilon}+e^{\gamma\epsilon}-2}{e^{-\gamma\epsilon}-1+\gamma\epsilon}\right), & h \geq \epsilon, \\ \dfrac{\gamma^2}{2}\left(\dfrac{-2e^{-\gamma h}+e^{-\gamma\epsilon}(e^{\gamma h}+e^{-\gamma h})}{e^{-\gamma\epsilon}-1+\gamma\epsilon}\right), & 0 \leq h < \epsilon. \end{cases} \tag{3.12}$$

The correlation function for the stationary process is even; therefore, using the classical result 3.2.4 it may be inferred that the SOU process is differentiable in the mean square sense. Hence, the SOU process is systematically different from the OU process, and only in the limit coincides with it.

Also, close to lag zero the correlation function of the OU process is concave upward (positive curvature) whereas the correlation of the SOU process is concave downward (negative curvature). This difference in curvature makes estimation of OU parameters γ, σ^2, based on observations from the SOU process a challenging task, which is addressed using an explicit subsampling scheme dictating an optimal choice of time lag.

More specifically, for a fixed positive $\varepsilon \ll 1$ the correlation function (3.10) of the SOU process for small lags $0 < h < \varepsilon$ may be approximated by $CF_{sou} \approx (1 - \text{cte}\ h^2)$, where the positive constant *cte* depends on γ and ε. More specifically, we obtain that the leading order terms for small $h < \varepsilon \ll 1$ are given by

$$CF_{SOU}(h) = 1 - \left(\frac{1}{\gamma\varepsilon} - \frac{1}{6} + O(\varepsilon)\right)\gamma^2\, h^2 + O(h^3). \tag{3.13}$$

If we attempt to fit the exponentially decaying function $CF_{ou} = e^{-\gamma h}$ to $(1 - \text{cte}\ h^2)$ at a fixed lag $\tau < \varepsilon$, then this is achieved for $\gamma \approx \text{cte}\ \tau$. Since γ depends on τ, this implies that there may not be a unique value of the parameter γ which may reproduce the correlation

function of the SOU process at every lag close to zero. Moreover, γ vanishes as the time lag $\tau \to 0$. Hence, the OU process does not reproduce the behavior of the SOU process at small lags. Our objective is to estimate γ and σ^2 by observing only the approximating SOU process Y_t^ε at discrete points, which may be realized by considering larger time lags.

The correlation function (3.10) associated to the SOU process for lags $h \geq \varepsilon$ verifies,

$$|CF_{sou}(h) - CF_{ou}(h)| \approx O(\varepsilon).$$

Therefore, it is expected that estimation of γ based on correlations of the SOU process at some lag $\tau > \varepsilon$ will have an error of the order (ε/τ). This intuitive explanation will be made rigorous in the ensuing sections in this chapter.

It is important to point out these features associated to the SOU process which distinguish the SOU process from the OU process. The behavior of the correlation function close to lag zero will play a major role in the estimation of the parameters of the OU process, and more general classes of random processes.

In our study, we are going to systematically address the optimality of covariance lags chosen for parametric estimation of the *unobserved* OU process based on the discrete data from the *observed* SOU process.

3.3 Fixed-rate Subsampling

Recall that now the *only available information* are $(N+1)$ *indirect* observations $U_n^\varepsilon = Y_{n\Delta}^\varepsilon$ extracted from the SOU process Y_t^ε by subsampling with a fixed time-step $\Delta > 0$.

The goal, nevertheless, is to estimate the parameters γ and σ^2 of the underlying OU process. We will study the estimators given by formulas (2.25) where we replace U_n by U_n^ε.

These approximate MLEs of γ and σ^2 are given by

$$\hat{\gamma}_\varepsilon = -\frac{1}{\Delta}\ln\left(\frac{\hat{r}_1^\epsilon}{\hat{r}_0^\epsilon}\right), \quad \hat{\sigma}_\varepsilon^2 = 2\hat{\gamma}_\varepsilon \hat{r}_0^\epsilon, \tag{3.14}$$

where

$$\hat{r}_k^\epsilon = \hat{r}_k^\epsilon(N,\Delta) = (1/N)\sum_{n=0}^{N-1} U_n^\varepsilon U_{n+k}^\varepsilon, \qquad \text{for } k = 0,1, \tag{3.15}$$

are the standard empirical estimators of the covariances $r_0^\varepsilon = K^\varepsilon(0)$, $r_1^\varepsilon = K^\varepsilon(\Delta)$ given by (3.4).

Consider the asymptotics of the estimators (3.14) under the *fixed-rate subsampling*, i.e., when $\varepsilon > 0$ is fixed, the time step length $\Delta > 0$ is fixed, and the number of observations $N \to \infty$. We will show that under fixed-rate subsampling the estimators are asymptotically biased. In the case of direct observability ($\varepsilon = 0$), the estimators of γ, σ^2 given by (3.14) are asymptotically consistent, and converge in L_2 to the corresponding true values with speeds of convergence proportional to $1/\sqrt{N}$.

First we compute the L_2-norms of the random variables $(\hat{r}_k^\epsilon - r_k^\varepsilon)$, where $\hat{r}_k^\epsilon$ are given by (3.15) and $r_k^\varepsilon = K^\varepsilon(k\Delta)$ given by (3.4). Since, the SOU process is a centered stationary Gaussian processs we may apply the result from proposition 2.4.3.

Proposition 3.3.1 (Asymptotics for "subsampled" covariance estimators) *Let $0 < \varepsilon \ll 1$ be fixed. Consider N discrete observations $U_n = Y_{n\Delta}^\varepsilon$ subsampled from the trajectory of an SOU process Y_t^ε at fixed uniform time step $\Delta > 0$. For $k = 0,1$, and fixed ε, Δ, consider the "subsampled" empirical covariance estimators $\hat{r}_k^\epsilon$,*

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

Let, for any non-negative integer k, $r_k^\varepsilon(\Delta) = K^\varepsilon(k\Delta) = E[U_n U_{n+k}]$, given by (3.4), denote the "true" covariances for the subsampled process U_n.

Then, as $N \to \infty$, the empirical covariance estimators $\hat{r}_k^\epsilon$ converge in L_2 to the covariances $r_k^\varepsilon(\Delta)$ associated to the "subsampled" SOU process $U_n = Y_{n\Delta}^\varepsilon$. Moreover, the L_2-speed of convergence of $\hat{r}_k^\epsilon$ to $r_k^\varepsilon(\Delta)$ is proportional to $1/\sqrt{N}$.

Proof. There are two possible scenarios, (i) $\Delta \geq \varepsilon$, and (ii) $0 < \Delta < \varepsilon$. Define the following scalars,

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

(i) For $\Delta \geq \varepsilon$, using (2.15) from proposition 2.4.3, we obtain, for each non-negative integer k,

$$J_k^\epsilon = Cov(\hat{r}_k^\epsilon, \hat{r}_k^\epsilon) = \frac{1}{N}\left(f(0) + 2\sum_{j=1}^{N-1} f(j)\right) - \frac{2}{N^2}\sum_{j=1}^{N-1} jf(j), \tag{3.16}$$

where $f(j) = f_{k,\varepsilon,\Delta}(j) = \left((r_k^\varepsilon)^2 + r_{j+k}^\varepsilon r_{j-k}^\varepsilon\right) = f(-j)$.

Then, the standard *empirical covariance estimators* $(\hat{r}_k^\epsilon)_{k=0,1}$ given by (3.15), are such that the L_2-norms $J_0^\epsilon = \|\hat{r}_0^\epsilon - r_0^\varepsilon\|_{L_2}^2$, and $J_1^\epsilon = \|\hat{r}_1^\epsilon - r_1^\varepsilon\|_{L_2}^2$ are given by,

$$\begin{aligned} J_0^\epsilon &= \frac{2C_1^2}{N}\left(\frac{C_0^2}{C_1^2} + \frac{2b^2}{1-b^2} - \frac{2b^2\left(1-b^{2N}\right)}{N(1-b^2)^2}\right), \\ J_1^\epsilon &= \frac{C_1^2}{N}\left(\frac{C_0^2}{C_1^2} + \frac{2C_0b^2}{C_1} + \frac{b^2(3+b^2)}{1-b^2} - \frac{2b^2B_1}{N}\right). \end{aligned} \tag{3.17}$$

where $B_1 = \left[(C_0/C_1) + \left(1 + 2b^2 - b^4 - 2b^{2N}\right) / \left(1-b^2\right)^2\right]$.

From (3.17) it may be inferred that $J_0^\epsilon \to 0$, and $J_1^\epsilon \to 0$, which is the required result.

(ii) Similarly, for $0 < \Delta < \varepsilon$, we use the expression (2.15) from proposition 2.4.3, and

arrive at the (3.16), i.e.,

$$J_k^\epsilon = Cov(\hat{r}_k^\epsilon, \hat{r}_k^\epsilon) = \frac{1}{N}\left(f(0) + 2\sum_{j=1}^{N-1} f(j)\right) - \frac{2}{N^2}\sum_{j=1}^{N-1} jf(j),$$

where $f(j) = f_{k,\varepsilon,\Delta}(j) = \left((r_k^\varepsilon)^2 + r_{j+k}^\varepsilon r_{j-k}^\varepsilon\right) = f(-j)$. For each $k = 0, 1$, since, $\Delta < \varepsilon$, the sums in J_k^ϵ can be decomposed into two parts, namely

$$\sum_{j=1}^{N-1} f(j) = \sum_{j=1}^{M} f(j) + \sum_{j=M+1}^{N-1} f(j), \quad \text{and} \quad \sum_{j=1}^{N-1} jf(j) = \sum_{j=1}^{M} jf(j) + \sum_{j=M+1}^{N-1} jf(j),$$

where M is a positive integer such that $M\Delta \leq \varepsilon < (M+1)\Delta$, and $N \gg M$. Then, for j between 1 and M, the value of $f(j)$ is determined by the covariances $K^\varepsilon(h)$, $h \leq \varepsilon$, given by (3.4). On the other hand, for $j > M$, $f(j)$ is determined by $K^\varepsilon(h)$, $h > \varepsilon$. By evaluating the sums with the appropriate expressions for the covariances $r_j^\varepsilon = K^\varepsilon(j\Delta)$, we get the required convergence for fixed ε, Δ and as $N \to \infty$.

■

As a result of the previous proposition we are able to conclude the following result concerning the asymptotics of the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$.

Proposition 3.3.2 (Asymptotic Bias of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$) *For fixed ε and Δ the following convergence holds in L_2 as $N \to \infty$, namely,*

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

where

$$G = -(1/\Delta)\ln\left(K^\varepsilon(\Delta)/K^\varepsilon(0)\right) \quad \textit{and} \quad S = 2GK^\varepsilon(0), \tag{3.18}$$

and where the covariances $K^\varepsilon(0)$ and $K^\varepsilon(\Delta)$ are given by (3.4).

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

$$Bias_\gamma = G - \gamma, \quad Bias_{\sigma^2} = S - \sigma^2. \tag{3.19}$$

The explicit expressions of these asymptotic biases are given below in (3.20) *and* (3.21).

Proof. Since, the SOU process Y^ϵ is, for a fixed ε, a stationary Gaussian process from which we subsample the observations U^ε_n with a fixed time-step Δ, therefore, the proof applies the exactly same generic principles as the proof of proposition 2.5.2 above, and we may directly apply the results from section 2.5 to the covariance estimators $\hat{r}^\epsilon_k$ for $k = 0, 1$ and to $\hat{\gamma}_\varepsilon = g\left(\hat{r}^\epsilon_0, \hat{r}^\epsilon_1\right)$ and $\hat{\sigma}^2_\varepsilon = s\left(\hat{r}^\epsilon_0, \hat{r}^\epsilon_1\right)$, given by (3.14). ■

As expected, *indirect estimation* of the OU process is less favorable than estimation based on *direct OU observations*, so that $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are not asymptotically consistent estimators, respectively, of the true OU parameters γ, σ^2, as $N \to \infty$, for a fixed value of ε and Δ. Instead, these estimators have non-zero asymptotic biases $(G - \gamma)$ and $(S - \sigma^2)$ given by (3.18), (3.19), that are functions of Δ, ε.

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

Proposition 3.3.3 (Favorable Regime for Consistency) *As seen in proposition 3.3.2, for fixed ε and Δ, the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ both have non-zero asymptotic biases $Bias_\gamma$ and $Bias_{\sigma^2}$ as $N \to \infty$, which depend only on $\varepsilon, \Delta, \gamma, \sigma$. Assume now that $\varepsilon \to 0$, and for each ε select a number $N = N(\varepsilon)$ of indirect observations of Y^ε_t and a subsampling rate $\Delta = \Delta(\varepsilon)$ such that $\Delta \to 0$ and $N\Delta \to \infty$.*

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

Proof. From formula (3.5), we see that as $\varepsilon \to 0$, we have $K^\varepsilon(0) \to \sigma^2/2\gamma$; then the expression of S given by (3.18) shows that whenever $G \to \gamma$ as $\varepsilon \to 0$, we must also have

$S \to \sigma^2$. Hence we only need to study the asymptotic behavior of $Bias_\gamma$. Note first that whenever $\Delta = \Delta(\varepsilon) \geq \varepsilon$ and $\varepsilon \to 0$, we have in view of (3.10) and (3.19),

$$Bias_\gamma = -(1/\Delta) \ln \left(\frac{e^{-\gamma\varepsilon} + e^{\gamma\varepsilon} - 2}{2\left(e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon\right)} \right) = -\frac{\gamma\varepsilon}{3\Delta} \left(1 + \frac{\gamma\varepsilon}{6} + O(\varepsilon^2) \right). \tag{3.20}$$

We have several cases to consider.

1. Assume that $(\Delta/\varepsilon) \to \infty$ as $\varepsilon \to 0$. Then, for ε small enough, we have $\Delta \geq \varepsilon$, and (3.20) proves that $Bias_\gamma \to 0$, as $\varepsilon \to 0$, and, hence, $Bias_{\sigma^2} \to 0$.

2. Let Δ/ε remain bounded as $\varepsilon \to 0$. Then, there exist a subsequence $\varepsilon \to 0$ such that $\Delta/\varepsilon \to L$ for some non-negative L.

   (a) If $L \geq 1$ then for ε small enough we have $\Delta \geq \varepsilon$ and hence, in view of (3.20), we have $Bias_\gamma \approx (-\gamma\varepsilon/(3\Delta))$ so that $Bias_\gamma$ of the subsequence tends to the non-zero limit $(-\gamma/(3L))$.

   (b) Assume that $L < 1$. Then for ε small enough we have $\Delta < \varepsilon$ and hence, in view of (3.10) and (3.19), we have

   $$Bias_\gamma = -(1/\Delta) \ln \left(\frac{2\gamma(\varepsilon - \Delta)e^{\gamma\Delta} + e^{-\gamma\varepsilon}(e^{2\gamma\Delta} + 1) - 2}{2\left(e^{-\gamma\varepsilon} - 1 + \gamma\varepsilon\right)} \right). \tag{3.21}$$

   Successive Taylor expansions with respect to Δ, and ε in (3.21), leads to the following limit, namely,

   $$Bias_\gamma \to (-\gamma)\left(1 - L + L^2/3\right), \qquad \text{when } \varepsilon \to 0. \tag{3.22}$$

   But the quadratic polynomial $\left(1 - L + L^2/3\right)$ remains strictly positive for $0 \leq L < 1$. Hence, $Bias_\gamma$ tends to a non-zero limit in case 2.(b).

■

Proposition 3.3.3 clearly defines a **favorable regime for adaptive subsampling**. We have seen that the asymptotic biases of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$, namely, $Bias_\gamma$ and $Bias_{\sigma^2}$, tend to 0 as $\varepsilon \to 0$ if and only if $(\Delta/\varepsilon) \to \infty$. This strongly indicates that optimal *adaptive subsampling* schemes from indirect observations based on Y^ε should provide $N = N(\varepsilon)$ observations $U^\varepsilon_n = Y^\varepsilon_{n\Delta}$ subsampled from Y^ε_t at time interval $\Delta = \Delta(\varepsilon)$, under the following set of simultaneous conditions,

$$\varepsilon \to 0, \quad \Delta \to 0, \quad \Delta/\varepsilon \to \infty, \quad N\Delta \to \infty. \tag{3.23}$$

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

Under *fixed rate subsampling*, applying the general results on the asymptotic properties of empirical covariance estimators based on the observations from a stationary Gaussian processes as described in [10, Chapter X], the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are asymptotically Gaussian, i.e., the random vector $\sqrt{N}\left(\hat{\gamma}_\varepsilon - G, \hat{\sigma}^2_\varepsilon - S\right)$ converges to a Gaussian distribution with mean zero, and covariance matrix dependent on the true parameters γ, σ^2, ε and Δ.

Since, in particular, for each fixed ε, $\Delta > 0$, as $N \to \infty$, the empirical covariance estimators $\hat{r}^\varepsilon_0$, $\hat{r}^\varepsilon_1$ are asymptotically Gaussian [10]. The estimators, using (3.14), are given by $\hat{\gamma}_\varepsilon = g\left(\hat{r}^\varepsilon_0, \hat{r}^\varepsilon_1\right)$ and $\hat{\sigma}^2_\varepsilon = s\left(\hat{r}^\varepsilon_0, \hat{r}^\varepsilon_1\right)$, such that g, s have continuous second-order partial derivatives in a neighborhood of the true values r_0, r_1. Therefore, as $N \to \infty$, for each fixed ε, $\Delta > 0$, the estimators $\hat{\gamma}_\varepsilon = g\left(\hat{r}^\varepsilon_0, \hat{r}^\varepsilon_1\right)$ and $\hat{\sigma}^2_\varepsilon = s\left(\hat{r}^\varepsilon_0, \hat{r}^\varepsilon_1\right)$ are asymptotically Gaussian [10].

We now study these estimators under the conditions (3.23) in detail.

3.4 Adaptive Subsampling

Here we are able to show that the drift and diffusion estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are asymptotically consistent in the limit of $\varepsilon \to 0$ under the conditions given by (3.23). These conditions identify explicit numerical schemes which we present below.

Proposition 3.4.1 (Asymptotics of the Covariances) *Consider an adaptive subsampling scheme, based on $N = N(\varepsilon)$ indirect observations extracted from Y^ε_t by subsampling with time steps $\Delta = \Delta(\varepsilon)$. Then, under the conditions (3.23), the L_2 norms of the variables $(\hat{r}^\epsilon_0 - r^\varepsilon_0)$, and $(\hat{r}^\epsilon_1 - r^\varepsilon_1)$ converge to 0 with speeds of convergence proportional to $1/\sqrt{N\Delta}$.*

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

Proof. Define $J^\epsilon_k = E\left[(\hat{r}^\epsilon_k - r^\varepsilon_k)^2\right]$ for $k = 0, 1$ computed explicitly for $\Delta > \varepsilon$, by using proposition 2.4.3. Let

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

Then, we have

$$\begin{aligned} J^\epsilon_0 &= \frac{2C_1^2}{N}\left(\frac{C_0^2}{C_1^2} + \frac{2b^2}{1-b^2} - \frac{2b^2\left(1-b^{2N}\right)}{N(1-b^2)^2}\right), \\ J^\epsilon_1 &= \frac{C_1^2}{N}\left(\frac{C_0^2}{C_1^2} + \frac{2C_0b^2}{C_1} + \frac{b^2(3+b^2)}{1-b^2} - \frac{2b^2B_1}{N}\right), \end{aligned} \tag{3.24}$$

where $B_1 = \left[(C_0/C_1) + \left(1 + 2b^2 - b^4 - 2b^{2N}\right) / \left(1-b^2\right)^2\right]$. From (3.24) we obtain bounds for J^ϵ_0 and J^ϵ_1 given, for $\gamma\Delta \ll 1$, by

$$J^\epsilon_0 \le \frac{2C_0^2}{N} + \frac{2C_1^2}{\gamma N\Delta}, \qquad \text{and} \qquad J^\epsilon_1 \le \frac{C_0^2}{N} + \frac{2C_0C_1e^{-\gamma\Delta}}{N} + \frac{2C_1^2}{\gamma N\Delta}.$$

These inequalities show that $J_k^\epsilon \to 0$ under the *adaptive subsampling scheme* defined in (3.23). The exact expressions in (3.24) gives, as $\varepsilon \to 0$,

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

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

Proposition 3.4.2 *For each N, Δ, ε, the random variables $Z_0 = Z_0(N, \Delta, \varepsilon)$ and $Z_1 = Z_1(N, \Delta, \varepsilon)$ defined by,*

$$Z_0 = \frac{(\hat{r}_0^\epsilon - r_0^\varepsilon)}{\sqrt{J_0^\epsilon}}, \qquad Z_1 = \frac{(\hat{r}_1^\epsilon - r_1^\varepsilon)}{\sqrt{J_1^\epsilon}}, \tag{3.25}$$

have mean 0, variance 1, and covariance $E[Z_0 Z_1] = J_{01}^\epsilon / \sqrt{J_0^\epsilon J_1^\epsilon}$, where $J_0^\epsilon, J_1^\epsilon, J_{01}^\epsilon$ are given by (3.24), (3.27).

Then, under the conditions,

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

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

$$\begin{aligned} \hat{r}_0^\epsilon &= r_0^\varepsilon + \frac{\sqrt{2}r_0}{\sqrt{\gamma N\Delta}} Z_0 + \frac{Z_0}{\sqrt{N\Delta}} \left(O(\varepsilon^2) + O(\Delta^2) + O\left(\frac{1}{N\Delta}\right)\right), \\ \hat{r}_1^\epsilon &= r_1^\varepsilon(\Delta) + \frac{\sqrt{2}r_0}{\sqrt{\gamma N\Delta}} Z_1 + \frac{Z_1}{\sqrt{N\Delta}} \left(O(\varepsilon^2) + O(\Delta^2) + O\left(\frac{1}{N\Delta}\right)\right), \end{aligned} \tag{3.26}$$

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

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

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

then, the exact expression for the covariance $J_{01}^{\epsilon} = E\left[(\hat{r}_0^{\epsilon} - r_0^{\varepsilon})\,(\hat{r}_1^{\epsilon} - r_1^{\varepsilon})\right]$, is given by

$$J_{01}^{\epsilon} = \frac{2C_1^2}{N}\left(\frac{2C_0 b}{C_1} + \frac{2b^3}{1-b^2} - \frac{B_2}{N}\right), \tag{3.27}$$

where

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

Using Taylor expansions we obtain the following approximations,

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

from which we can deduce (3.26). ■

The following theorem presents the key results of our study.

Theorem 1 (Asymptotic Consistency of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$) *Consider an adaptive subsampling scheme, based on $N = N(\varepsilon)$ indirect observations extracted from Y_t^{ε} by subsampling with time steps $\Delta = \Delta(\varepsilon)$. Let the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ of γ, σ^2, be given by* (3.14)*. Then, under the following conditions,*

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad \Delta/\varepsilon \to \infty, \tag{3.28}$$

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

Moreover, the expected L_2-speed of convergence, proportional to $1/\sqrt{N\Delta}$, is achievable

under the following conditions which are stronger than (3.28)*,*

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta \to \infty, \quad N\varepsilon^2/\Delta < cte. \tag{3.29}$$

In particular, under stronger conditions than (3.29)*,* (3.28)*, namely,*

$$\varepsilon \to 0, \quad \Delta \to 0, \quad N\Delta^2 \to \infty, \quad N\varepsilon^2/\Delta^2 \to 0, \tag{3.30}$$

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

Proof. Substitute the expressions for empirical covariance estimators $\hat{r}_k^\epsilon$, given by (3.25), in the expressions for the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ defined in (3.14). In particular, the drift estimator $\hat{\gamma}_\varepsilon$ is given by,

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

Then, using Taylor expansions as $\varepsilon \to 0$ and using arguments similar to those given in the proof of proposition 2.6.4, we obtain the following first-order L_2-approximation for $\hat{\gamma}_\varepsilon$ given by,

$$\hat{\gamma}_\varepsilon = \gamma - \frac{\gamma\varepsilon}{3\Delta} - Z_\gamma + R_\gamma \times O\left(\frac{1}{N\Delta}\right) + \frac{\varepsilon}{\Delta} \times O(\varepsilon), \tag{3.31}$$

where the zero mean random variable Z_γ is given by,

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

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

$$\|Z_\gamma\|^2_{L^2} = \frac{2\gamma}{N\Delta}\left(1 + 3\gamma\Delta - \frac{1+O(\Delta)}{2\gamma N\Delta} + O\left(\frac{\varepsilon}{\Delta}\right) + O(\Delta^2) + O(\varepsilon)\right). \tag{3.32}$$

The remainder term $R_\gamma = R_\gamma(\Delta, \varepsilon, N)$ is uniformly bounded in L_2 norm. Therefore, using (3.31) and (3.32), under the conditions (3.28) the estimator $\hat{\gamma}_\varepsilon$ converges in L_2 to γ.

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

$$\sqrt{N\Delta}\,(\hat{\gamma}_\varepsilon - \gamma) = -\sqrt{N\Delta}Z_\gamma - \frac{\gamma\varepsilon\sqrt{N}}{3\sqrt{\Delta}} + R_\gamma \times O\left(\frac{1}{\sqrt{N\Delta}}\right). \tag{3.33}$$

Using (3.32), (3.33) we see that the L^2-norm of $\sqrt{N\Delta}\,(\hat{\gamma}_\varepsilon - \gamma)$ converges to a constant under conditions (3.29). Under the *adaptive subsampling scheme* (3.29), we assume $N\varepsilon^2/\Delta \to 0$ to deduce that the asymptotic variance of estimation errors converge to the same constant as in the case of *direct estimation* (see proposition 2.6.4), i.e.,

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

Similarly, given the conditions (3.28), the diffusion estimator $\hat{\sigma}^2_\varepsilon = 2\hat{\gamma}_\varepsilon \hat{r}^\epsilon_0$ converges in L_2 to the true value σ^2, and, hence, is asymptotically consistent. Furthermore, under conditions (3.30) we obtain,

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

■

The main conclusion of theorem 1 is that under conditions (3.29) the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$, based on indirect estimation, are *asymptotically consistent* estimators of γ, σ^2, with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta}$.

A natural objective is to optimally select $\Delta = \Delta(\varepsilon)$ and $N = N(\varepsilon)$, verifying the conditions (3.29), in order to achieve the fastest speed of convergence. A *pragmatic interpretation* of the conditions (3.29) is that, as $\varepsilon \to 0$, one selects $\Delta = \Delta(\varepsilon)$ such that

$$\Delta \to 0, \quad \Delta \gg \varepsilon, \quad \text{and N verifies, } (1/\Delta) \ll N < cte(\Delta/\varepsilon^2). \tag{3.34}$$

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

$$cte\left(\frac{\varepsilon}{\Delta}\right) < \frac{1}{\sqrt{N\Delta}} \ll 1. \tag{3.35}$$

Clearly, the lower bound (ε/Δ) in (3.35) is the best L_2-speed of convergence achievable under the conditions (3.29). This speed is attained when $N \sim \Delta/\varepsilon^2 \to \infty$, which corresponds to a global time interval of observations $T^* = N\Delta = cte(\Delta^2/\varepsilon^2)$.

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

In the following corollary we provide a particular example of the optimal criterion identified by the pragmatic interpretation (3.34).

Corollary 3.4.3 (Power Law Criterion for Optimal Subsampling) *As $\varepsilon \to 0$, assume that $N(\varepsilon)$ and $\Delta(\varepsilon)$ are given by powers of ε, namely, $N(\varepsilon) = \varepsilon^{-\eta}$, $\Delta(\varepsilon) = \varepsilon^{\alpha}$. Then,*

1. *as $\varepsilon \to 0$, for any α, η such that $\alpha \in (0,1)$, $\eta > \alpha$, the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are asymptotically consistent in L_2 norm.*

2. *Moreover, as $\varepsilon \to 0$, under stronger conditions, namely, for any α, η such that*

$\alpha \in (0,1)$, $\alpha < \eta \leq 2-\alpha$, the estimators converge with an L_2-speed of convergence proportional to $1/\sqrt{N\Delta} = \varepsilon^{(\eta-\alpha)/2}$.

3. *The best speeds of convergence are reached when $\alpha > 0$ is close to 0, and $\eta = 2-\alpha$. Then, we obtain $\Delta = \varepsilon^{\alpha}$, $N = \varepsilon^{-(2-\alpha)}$, and the global time of observations $N\Delta = \varepsilon^{-2(1-\alpha)}$.*

3.5 Numerical Results

We now study numerically a few typical examples of adaptive subsampling schemes ensuring asymptotic consistency of estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$. In view of the corollary 3.4.3, we let $\Delta(\varepsilon) = \varepsilon^\alpha$ where $\alpha \in (0,1)$, and the number of observations $N \gg (\Delta/\varepsilon^2)$. The following numerical results show that as $\varepsilon \to 0$, $Bias_\gamma(\Delta,\varepsilon)$ and $Bias_{\sigma^2}(\Delta,\varepsilon)$ converge to 0 if and only if $(\Delta(\varepsilon)/\varepsilon) \to \infty$ (See proposition 3.3.3). As evident in the following numerical study and from corollary 3.4.3 for smaller values of $\alpha \in (0,1)$, the convergence of the $Bias_\gamma$ and $Bias_{\sigma^2}$ to zero is faster.

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

The underlying discretized trajectory of X_t is generated using a hybrid of Euler-Maruyama and second-order Runge-Kutta discretization schemes for the SDE (3.1), with a time-step length of $d = 10^{-4}$ and total time interval $T = 900$, thus providing 9×10^6 points of OU-trajectory. To generate SOU observations, we average the simulated OU observations over

a sliding window of length ε, for the following values of ε,

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

We consider 3 examples of *adaptive subsampling* schemes, namely, when observations are subsampled with time-step $\Delta(\varepsilon) = \varepsilon^{0.5}$, $\Delta(\varepsilon) = \varepsilon$, and $\Delta(\varepsilon) = \varepsilon^2$. In each one of these 3 cases, for each simulated trajectory of the SOU process, we compute the subsampled estimators $\hat{\gamma}_N$ and $\hat{\sigma}^2_N$ given by (3.14). Figure 3.1 shows numerical verification of the consistency results obtained in corollary 3.4.3. Errors (in %) in the figure is defined to be the absolute value of the relative bias in the estimates. For instance, for the error in the estimation of γ, we have

$$Error = \left| \frac{\hat{\gamma}_\varepsilon(N, \Delta) - \gamma}{\gamma} \right|.$$

1. **Case** $\Delta(\varepsilon) = \varepsilon^{0.5}$: Results are displayed in the top part of Figure 3.1. The empirical relative bias (errors) of subsampled estimators tend to zero as $\varepsilon \to 0$, as expected, since $\Delta(\varepsilon)/\varepsilon \to \infty$ in this case.

2. **Case** $\Delta(\varepsilon) = \varepsilon$: Results are displayed in the middle part of Figure 3.1. The empirical relative bias (errors) of the subsampled estimators converge to a non-zero value, as $\varepsilon \to 0$, as expected, since $\Delta(\varepsilon)/\varepsilon$, is bounded in this case.

   Formula (3.20) for the asymptotic bias give $Bias_\gamma \approx -\gamma/3$ and $Bias_{\sigma^2} \approx -\sigma^2/3$, which fit very well with the numerical results.

3. **Case** $\Delta(\varepsilon) = \varepsilon^2$: Results are displayed in the bottom part of Figure 3.1. The empirical relative bias (errors) of the subsampled estimators increase as $\varepsilon \to 0$, as expected, since $\Delta(\varepsilon)/\varepsilon \to 0$ in this case.

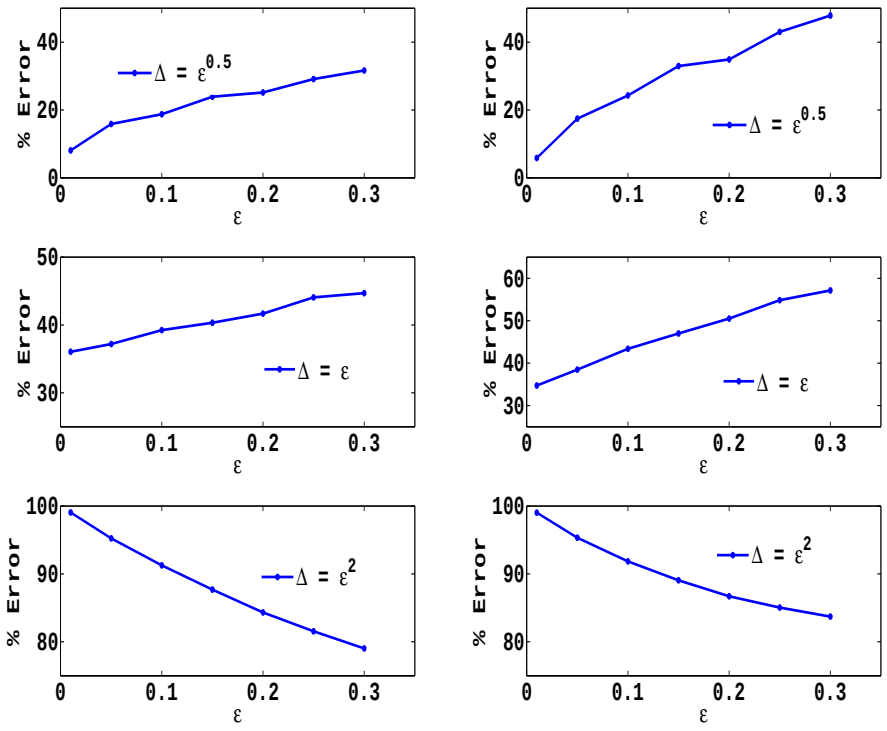

Figure 3.1: Relative (%) errors in $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ based on observations from the SOU process subsampled with three different strategies. Left part - Errors in $\hat{\gamma}_\varepsilon$, Right part - Errors in $\hat{\sigma}^2_\varepsilon$. Top part - Subsampling with $\Delta = \varepsilon^{0.5}$: errors converge to 0 with speed of convergence proportional to $\varepsilon^{0.5}$. Middle part - Subsampling with $\Delta = \varepsilon$: errors converge to a constant ($\approx 33\%$) with speeds of convergence proportional to ε. Bottom part - Subsampling with $\Delta = \varepsilon^2$: errors increase to 100%.

3.6 Concluding Remarks

Our objective, aforementioned, is to consistently estimate parameters of an underlying process, assumed to be unobservable, based on observations subsampled from an observable approximating process. This objective has been explicitly illustrated with the above prototypical example, where the unobservable process is taken to be the stationary *Ornstein-Uhlenbeck* (denoted as OU) process (3.1) characterized by unknown parameters, drift $\gamma > 0$ and diffusion σ^2. These parameters are estimated given discrete set of observations from an observable process, taken to be the *Smoothed Ornstein-Uhlenbeck* (denoted by SOU) process defined in (3.2). The SOU process, in particular, is obtained by locally averaging the OU process over a moving window of length $\varepsilon > 0$. This local averaging makes the OU SDE (3.1) a misfit to the data sampled from the SOU process with very high frequency. Estimation of parameters under such mismatch (which we refer to as *indirect observability* of the unobserved process) is the central focus of this thesis, and this particular example (OU/SOU) effectively presents the main idea, along with the exact conditions under which the estimation is not only possible, but consistent.

In this first part of the thesis, by considering the SOU process, we have restricted ourselves to the Gaussian framework, where exact computations of the errors in the estimation were possible. The expressions for the errors derived, thus, lead to characterization of explicit conditions to guarantee consistency of the parameter estimators. For instance, given $N = N(\varepsilon)$ discrete observations subsampled from the SOU process Y_t^ε with uniform time step $\Delta = \Delta(\varepsilon)$, then, under the conditions,

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

the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ converge in L_2 to the true unknown values γ, σ^2 corresponding to the OU process. The numerical study (figure 3.1) presented above, verifies these results.

Although the example of the SOU process is specific, it throws light on how *indirect observability* manifests itself in estimation problems. This gives us strong pointers for estimation under indirect observability in more generic examples.

In the second part of the thesis we apply some of the results to more complex examples, particularly, to data obtained from multiscale dynamics.

Chapter 4

Application to Multiscale Dynamics

4.1 Introduction

In chapter 3 we presented a favorable situation, where the *unobserved* OU process X_t is approximated by *observed* Gaussian SOU processes Y_t^ε, to characterize explicitly the family of optimal subsampling regimes leading to consistent estimators having the best L_2-speeds of convergence under *indirect observability*. This specific framework replicates the scenario observed in several applications where a mismatch between the data and the stochastic model impedes the estimation procedure, so that appropriate adaptive subsampling schemes become necessary to obtain consistent estimates, and good L_2-speeds of convergence.

In many applications, it is desirable to capture the *stylized statistical features* of high-dimensional multiscale dynamics by fitting a reduced (single-scale) system of SDEs to the observed dataset. Although the reduced (or limiting) equation may capture the important statistical features of the full system at large scales, there may be a mismatch between the reduced system and full system at small scales. Mathematically this is explained as follows.

Consider multiscale dynamics of $\vec{w}_\varepsilon(t) = (\vec{u}_\varepsilon(t), \vec{v}_\varepsilon(t))$, where $\vec{u}_\varepsilon(t)$ form the slow variables (resolved part), i.e., they evolve on a longer time scale, say, of $O(1)$ or $O(\varepsilon^{-1})$; and $\vec{v}_\varepsilon(t)$ are the fast variables (unresolved part), i.e., as compared to $\vec{u}_\varepsilon$ they evolve on a shorter time scale, say, of order $O(\varepsilon)$, for $\varepsilon \ll 1$. Then, for specific dynamics under some conditions, in the limit of infinite scale separation, i.e., as $\varepsilon \to 0$, the slow variables $\vec{u}_\varepsilon$ converge to a process $\vec{U} = \{U(t)\}$ in an appropriate sense. The single-scale reduced model U, then, reproduces the behavior of the slow variables $\vec{u}_\varepsilon$, for each $\varepsilon > 0$, only on larger time-scales. Multiscale systems with the limiting model given by a Markov process have been studied extensively in the past [14, 33, 35, 56, 57, 71, 75, 77, 36].

In an attempt to identify the limiting process $\vec{U}$, given the high frequency observations from the resolved part $\vec{u}_\varepsilon$ of the full system, one fails to infer accurately the values of parameters associated to $\vec{U}$, precisely, due to the mismatch at small time scales. Subsampling the given observations from the full system leads to improvement in the accuracy of the estimates. Nevertheless, questions pertaining to optimal subsampling frequency, and asymptotic properties of the estimators (maximum likelihood estimators, etc.) are subtle, and have become the focus of research. This estimation problem has been discussed in the context of averaging and homogenization of multiscale diffusions in [72, 73, 76] (such that limiting model is a system of SDEs). In [73, 76] the estimation of homogenized multiscale diffusions is studied using likelihood methods, where as [72] presents an "equation free" modeling procedure. Non-parametric estimation of diffusions from the multiscale time series using spectral properties of the generator is presented in [26].

In particular, we consider two examples where the objective is to estimate parameters of the limiting equations corresponding to multiscale dynamics. The limiting equation is obtained using the method of homogenization (see appendix 0.15 for an example).

In many situations the dynamics of the full multiscale system - and consequentially those of the limiting reduced system - is not known. For instance, in the case of multiscale diffu-

sions, the drift and diffusion coefficients of the full system and thus the reduced system may depend upon unknown parameters, or the fast variables may be too complicated. Therefore, it is critical to develop statistical techniques to estimate parameters of the limiting reduced equation given only the observations from the slow variables.

Estimation of the limiting equation associated to a multiscale system naturally falls in the framework of *indirect observability*, which refers to estimation of an underlying unobserved process X_t based on observations from an approximating process Y_t^ε. The slow variables $\vec{u}_\varepsilon$ in the full system correspond to the observed approximating process Y_t^ε, and the limiting model U naturally takes the role of the *unobserved* process X_t.

The example of OU/SOU process presented in chapter 3 was suited for explicit computations of the errors in the estimation, which were then utilized to characterize the exact optimal subsampling schemes, i.e., ensuring asymptotic consistency of the estimators, c.f., (3.34). On the contrary, explicit computations are difficult for the multiscale models considered here. Therefore, we work with the intuition gained from the exact study of the OU/SOU example and conjecture our results for the multiscale diffusions. We support our conjecture with convincing numerical results and present arguments to demystify the observed behavior of the estimators.

In generic cases, the adequate convergence speed of good adaptive subsampling schemes can be identified by expanding the correlation function of Y_t^ε for small lags, or alternatively by the L_p-speed of convergence of the approximating process Y_t^ε to X_t, as $\varepsilon \to 0$. This approach, used in chapter 3, is extended in chapter 6 to more general class of stochastic models under indirect observability. The optimal subsampling schemes, for instance, for the Additive triad model presented below are characterized by, the scaling of the correlation function around lag zero, and the local behavior of the correlations as presented above in section 3.2.

4.2 Additive Triad

The first model which we consider is the additive triad model [64]

$$\begin{aligned} dx_t &= \frac{1}{\varepsilon}A_1 y_t z_t dt, \\ dy_t &= \frac{1}{\varepsilon}A_2 x_t z_t dt - \frac{1}{\varepsilon^2}g_2 y_t dt + \frac{1}{\varepsilon}s_2 dW_1(t), \\ dz_t &= \frac{1}{\varepsilon}A_3 x_t y_t dt - \frac{1}{\varepsilon^2}g_3 z_t dt + \frac{1}{\varepsilon}s_3 dW_2(t), \end{aligned} \tag{4.1}$$

where $A_1 + A_2 + A_3 = 0$, g_i, s_i are known positive parameters, W_1, W_2 are standard Brownian motions, and $\varepsilon > 0$ is the scale separation parameter. The stationary covariance of x_t can be computed explicitly for small lags and is given by

$$\frac{E[x_t x_{t+h}]}{E[x_t^2]} \approx 1 - C\frac{h^2}{\varepsilon^2}, \tag{4.2}$$

where $C = \gamma(g_2 + g_3)/2$ with γ in (4.4). Details of the derivation are provided in section 4.3.

It has been shown that x_t converges weakly to the OU process X_t as $\varepsilon \to 0$ [64]. The limiting equations are obtained by homogenization (see appendix 0.15) and are given by

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{4.3}$$

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

$$\gamma = \frac{-A_1}{2(g_2 + g_3)}\left(\frac{A_2 s_3^2}{g_3} + \frac{A_3 s_2^2}{g_2}\right), \qquad \sigma^2 = \frac{(A_1 s_2 s_3)^2}{2g_2 g_3 (g_2 + g_3)}. \tag{4.4}$$

The convergence of $x(t)$ in (4.1) to the OU process can be proved using the homogenization

procedure [64, 77]. The derivation utilizes the asymptotic expansion of the backward Kolmogorov equation. The weak convergence implies convergence of expectations of bounded functions. Nevertheless, we conjecture that moments of x_t converge to the moments of the Ornstein-Uhlenbeck process X_t in (4.3) as $\varepsilon \to 0$. Numerical results confirming the convergence of the correlation function and the (*generalized*) kurtosis (fourth-order moment measuring departures from Gaussianity) given as,

$$Kurt(\tau) = \frac{\langle x^2(t)x^2(t+\tau)\rangle_t}{\langle x^2(t)\rangle^2 + 2\langle x(t)x(t+\tau)\rangle_t^2}, \tag{4.5}$$

as $\varepsilon \to 0$ are depicted in Figure 4.1. Here $< . >_t$ denotes time averaging, which due to ergodicity [59] of the process, is equivalent to taking average with respect to the invariant density computed below in proposition 4.3.1. Note that in the benchmark case of a stationary Gaussian process $x(t)$, the value of $Kurt(\tau) = 1$.

Thus, we expect that for small enough ε the data generated by x_t in the triad model is close to the OU process in statistical sense. In particular, the correlation function of x_t converges to the exponential with the exponent γ in (4.4). Therefore, the estimation procedure should yield parameter values which are close to the analytical asymptotic expressions in (4.4). So, we use expressions (4.4) to test the performance of the parametric estimation under indirect observability when the data are generated by x_t in the triad in (4.1).

Subsampling for the Additive Triad Model

Now we test various subsampling strategies for adaptive parametric estimation of the underlying OU- process X_t, when the estimators are computed from indirect observations, namely the x_t data generated by the triad model. For multiple values of ε, we focus on the natural OU-process parameter estimators (3.14), but computed with datasets $U_n = x_{n\Delta(\varepsilon)}$

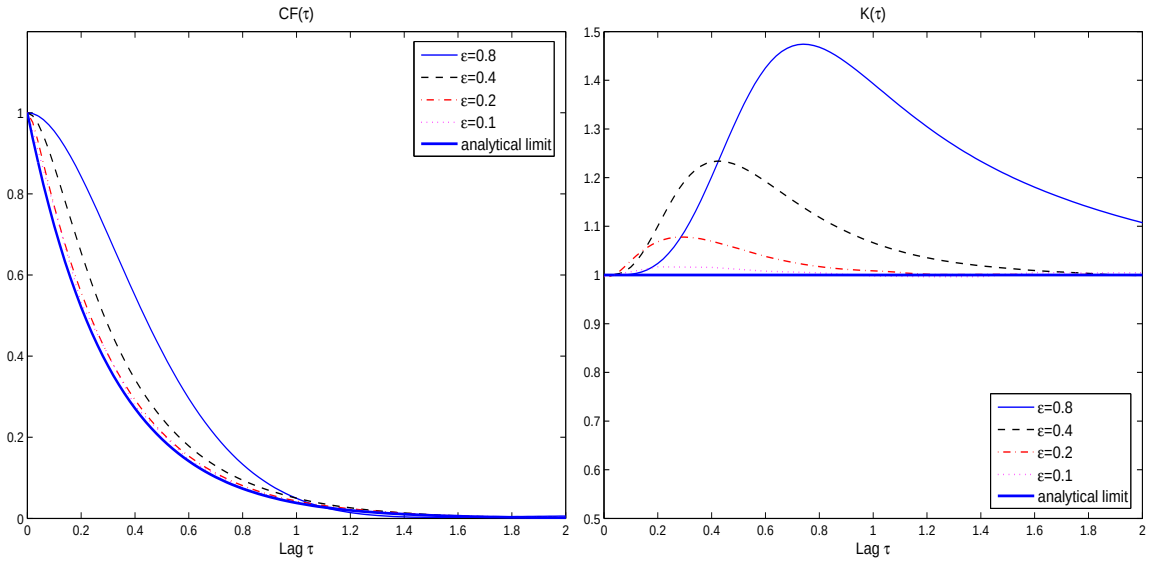

Figure 4.1: Left part - Convergence, as $\varepsilon \to 0$, of stationary correlation function $E[x_t x_{t+\tau}]$ in the Triad model (4.1) to correlation function $e^{-\gamma\tau}$ of the effective equation. Note that the graphs for $\varepsilon = 0.1$ and $\varepsilon = 0.2$ nearly overlap, and thus, are not distinctively visible in the figure. Right part - Stationary kurtosis (4.5) in the triad model (4.1) measuring departure, as $\varepsilon \to 0$, from Gaussianity ($Kurt(\tau) = 1$).

generated by numerical simulations of the triad model. As $\varepsilon \to 0$, we then compare the behavior of these estimators with the desired true parameter values , given by the explicit expressions (4.4). The triad model parameters in (4.1) 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

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

With this choice of triad parameters values, the drift and diffusion coefficients of the unobservable limit process X_t are computed by formula (4.4), which yields the values

$$\gamma = 3.2625, \quad \sigma = 6.75. \tag{4.6}$$

To determine the appropriate subsampling strategy for the dataset generated by the triad equations we compare the correlation functions of the SOU Process Y_t^ε given by (3.13) and of x_t in (4.2), 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 h, the respective correlations arr given by

$$CF_{\{Y_t^\varepsilon\}}(h) \sim 1 - \text{cte}\, h^2/\varepsilon, \quad CF_{\{x_t\}}(h) \sim 1 - \text{cte}\, h^2/\varepsilon^2,$$

where the generic notation "cte" denotes various numerical constants. Therefore, ε^2 in the triad model plays the role of ε in the SOU process, and when N, Δ are powers of ε the necessary condition for the consistency of the subsampling strategy for the triad model should be given by

$$\Delta = \varepsilon^{2\alpha}, \quad \alpha \in (0,1), \qquad N = \varepsilon^{-2\beta}, \quad \alpha < \beta.$$

Moreover, since we conjectured that ε^2 plays the role of ε in the SOU process, then when N is large enough (i.e. $N \gg \Delta/\varepsilon^4$) the bias for the adaptive indirect parametric estimation from the triad data is proportional to $\gamma\varepsilon^2/\Delta$, i.e.

$$\hat{\gamma}_\varepsilon - \gamma \sim \text{cte}\, \frac{\gamma\varepsilon^2}{\Delta} \quad \text{for} \quad N \gg \frac{\Delta}{\varepsilon^4}. \tag{4.7}$$

The size of datasets is chosen to be $N = 500,000$ and is kept constant in all simulations; $N \gg \Delta/\varepsilon^4$ holds for the smallest $\varepsilon = 0.1$ and largest $\Delta = 0.1$ considered here. Therefore,

the term on the right-hand side of (4.7) is the dominant term in the bias.

To support the above conjecture we compare three subsampling strategies. In particular, behavior of estimators (3.14) computed from the triad data with three subsampling strategies $\Delta = \varepsilon$, $\Delta = 4\varepsilon^2$, and $\Delta = \varepsilon^3$ is presented in Figure 4.2.

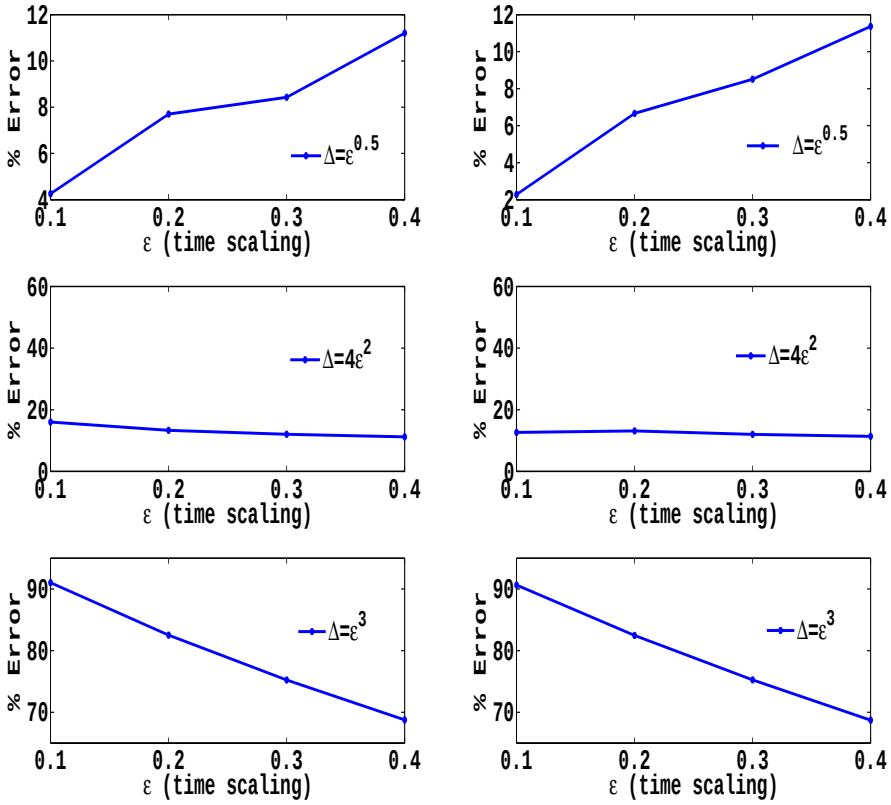

Figure 4.2: Relative (%) errors in $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ based on observations from the triad model (4.1) subsampled with three different strategies. Left part - Errors in $\hat{\gamma}_\varepsilon$, Right part - Errors in $\hat{\sigma}^2_\varepsilon$. Top part - Subsampling with $\Delta = \varepsilon^{0.5}$: errors vanish as $\varepsilon \to 0$. Middle part - Subsampling with $\Delta = 4\varepsilon^2$: errors converge to a constant as $\varepsilon \to 0$. Bottom part - Subsampling with $\Delta = \varepsilon^3$: errors increase to 100% as $\varepsilon \to 0$.

The top part of the Figure 4.2 demonstrates that when $\left(\Delta(\varepsilon)/\varepsilon^2\right) \to \infty$ estimates for γ and σ^2 are consistent with respect to the theoretical results in (4.4). On the other hand, errors remain bounded away from zero for the adaptive subsampling strategy such that $\left(\Delta(\varepsilon)/\varepsilon^2\right)$ is bounded. This is depicted in the middle part of Figure 4.2 where $\left(\Delta(\varepsilon)/\varepsilon^2\right)$ tends to a non-zero value, and hence, the errors converge to a constant strictly greater than zero. The bottom part of Figure 4.2 is based on subsampling scheme such that $\left(\Delta(\varepsilon)/\varepsilon^2\right) \to 0$, and the corresponding estimation errors increase to 100%. This provides strong evidence to support our conjecture, that the optimal subsampling strategy $\Delta \gg \varepsilon^2$ is the favorable subsampling regime for the estimation of the OU parameters from the triad data. The nature of the results is similar to the ones obtained for the SOU process.

The next section presents, in detail, the derivation of the approximation to the correlation function associated to the slow variable in the triad system (4.1). The scaling of the correlation function for lags close to zero is used to present the conjectured optimal adaptive subsampling scheme for consistent estimators.

4.3 Local Correlations to Identify Subsampling Schemes

The correlation function for the slow variable x_t in the additive triad system is derived as follows.

Proposition 4.3.1 (Invariant Density) *The Additive triad system* (4.1) *is asymptotically stationary with an invariant density given by,*

$$\Pi(x, y, z) = K \exp\left(-C_1 x^2 - C_2 y^2 - C_3 z^2\right), \tag{4.8}$$

where

$$C_1 = -A_1^{-1}\left(\frac{A_2 g_2}{s_2^2} + \frac{A_3 g_3}{s_3^2}\right), \qquad C_2 = \frac{g_2}{s_2^2}, \qquad C_2 = \frac{g_3}{s_3^2}, \tag{4.9}$$

and K is a normalizing constant.

Proof. The Fokker-Planck equation (see appendix 0.13) associated to the system of SDEs (4.1) is given as,

$$\begin{aligned}\partial_t P = &-\frac{1}{\varepsilon}\left(A_1\partial_x(yzP) + A_2\partial_y(xzP) + A_3\partial_z(xyP)\right)\\ &+\frac{1}{\varepsilon^2}\left(g_2\partial_y(yP) + \frac{1}{2}s_2^2\partial_{yy}^2(P) + g_3\partial_z(zP) + \frac{1}{2}s_3^2\partial_{zz}^2(P)\right).\end{aligned} \tag{4.10}$$

where the probability density $P = P(t, x, y, z \mid x_0, y_0, z_0)$. The expression in (4.10) may be represented by,

$$\partial_t P = \left(\frac{1}{\varepsilon}\mathcal{L}_{nl} + \frac{1}{\varepsilon^2}\mathcal{L}_{ou}\right)P,$$

where $\mathcal{L}_{nl}$ is the differential operator corresponding to the nonlinear terms and $\mathcal{L}_{ou}$ corresponds to the OU terms. The invariant density $\Pi = \Pi(x, y, z) = \lim_{t\to\infty} P(t, x, y, z \mid x_0, y_0, z_0)$, is characterized by,

$$\left(\frac{1}{\varepsilon}\mathcal{L}_{nl} + \frac{1}{\varepsilon^2}\mathcal{L}_{ou}\right)\Pi = 0.$$

The differential operator $\mathcal{L}_{ou}$ annihilates the bivariate Gaussian density function given by

$$\Pi_{ou}(y, z) = K_1 \exp\left(-(g_2y^2/s_2^2) - (g_3z^2/s_3^2)\right).$$

Under the energy conservation assumption ($A_1 + A_2 + A_3 = 0$) we have that L_{nl} annihilates

$$\Pi_{nl}(x, y, z) = K_2 \exp\left(-C_1x^2 - C_2y^2 - C_3z^2\right),$$

where C_1, C_2, and C_3 are any arbitrary non-negative constants. After substituting the product of the two densities we obtain,

$$\Pi(x, y, z) = K \exp\left(-C_1 x^2 - C_2 y^2 - C_3 z^2\right),$$

where C_1, C_2, and C_3 are as given by (4.9). ■

Therefore, the vector process (x_t, y_t, z_t) is asymptotically multivariate Gaussian with zero mean, and the covariance given by a (3×3) diagonal matrix with diagonal entries determined by $(1/2C_1)$, $(1/2C_2)$, and $(1/2C_3)$ (4.9).

Proposition 4.3.2 (Stationary Correlation Function) *The correlation function $CF_{atd}(h)$ associated to the slow variable x_t in the additive triad system* (4.1)*, is given by*

$$CF_{atd}(h) = 1 - \left(\frac{\gamma(g_2 + g_3)}{2\varepsilon^2}\right) h^2 + O(h^3), \tag{4.11}$$

where γ is the drift coefficient given by (4.4)*, and $h < \varepsilon^2 \ll 1$ such that ε^2 is the time scale of the fast variables.*

Proof. The dynamics of slow variable x_t is given by,

$$dx_t = A_1 y_t z_t \frac{dt}{\varepsilon}.$$

Then, integrating above equation over a small interval $[t, t + h]$ gives

$$x_{t+h} - x_t = \int_t^{t+h} A_1 y_s z_s \frac{ds}{\varepsilon}. \tag{4.12}$$

Now by the Itô formula, the product process $(y_t\ z_t)$, for $s > t$ is given by,

$$y_s\ z_s = y_t\ z_t + \int_t^s x_u \left(A_2 z_u^2 + A_3 y_u^2\right) \frac{du}{\varepsilon} - \int_t^s (g_2 + g_3) y_u z_u \frac{du}{\varepsilon^2} \ldots + \int_t^s s_2 z_u \frac{dW_1(u)}{\varepsilon} + \int_t^s s_3 y_u \frac{dW_2(u)}{\varepsilon}. \tag{4.13}$$

Substitute (4.13) in (4.12), and retain $O(h)$ terms to obtain the Euler-Maruyama discretization for the x-dynamics given by

$$x_{t+h} = x_t + A_1 y_t z_t \frac{h}{\varepsilon}, \qquad \text{for } h < \varepsilon^2 \ll 1. \tag{4.14}$$

Multiplying both sides of (4.14) by x_t and taking expectations with respect to the invariant density P (4.8), we obtain

$$E[x_{t+h} x_t] = E[x_t^2] + A_1 E[x_t y_t z_t] \frac{h}{\varepsilon} = E[x_t^2].$$

This shows that the correlation function of x is close to 1, for $h < \varepsilon^2 \ll 1$, which is a redundant approximation.

Hence, using Itô-Taylor expansion [58] we derive a higher-order discretization scheme for the x-dynamics, namely,

$$x_{t+h} = x_t + \frac{A_1 y_t z_t}{\varepsilon} h + \frac{A_1}{2\varepsilon^2} x_t (y_t^2 A_3 + z_t^2 A_2) h^2 \ldots - \frac{A_1 y_t z_t}{2\varepsilon^2} (g_2 + g_3) h^2. \tag{4.15}$$

Multiplying (4.15) by x_t and taking expectations with respect to the invariant density (4.8) gives

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

where $E[x_t^2] = (2C_1)^{-1}$, with C_1 given by (4.9).

Hence, the stationary correlation function $CF_{atd}(h)$ associated to the slow variable x_t in the additive triad system is given by,

$$CF_{atd}(h) = 1 - \left(\frac{\gamma(g_2 + g_3)}{2\varepsilon^2}\right) h^2 + O(h^3), \qquad \text{for } h < \varepsilon^2 \ll 1,$$

where γ is the drift coefficient (4.4) of the limiting process X_t (4.3). Note that as $\varepsilon \to 0$, for $h = h(\varepsilon)$ and $h \leq \varepsilon^2$, the correlation function $CF_{atd}(h(\varepsilon)) \to CF_{atd}(0) = 1$. ■

Comparing the stationary correlation function $CF_{atd}(h)$ (4.11) to the correlation function $CF_{sou}(h)$ (3.13) associated to the SOU process, we conjecture that for estimation of unknown parameters in (4.3), $\Delta \gg \varepsilon^2$ is the correct critical subsampling when observations come from the triad model in (4.1). Moreover, the analogous *adaptive subsampling scheme* to ensure consistency of estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ based on the data from the triad is determined by,

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

4.4 Truncated Burgers-Hopf Model

In this section we consider the subsampling problem for the Truncated Burgers-Hopf (TBH) model

$$\partial_t U_\Lambda(x,t) + \frac{1}{2}\partial_x P_\Lambda U_\Lambda^2(x,t) = 0, \tag{4.16}$$

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

$$P_\Lambda u(x,t) = \sum_{|k|\leq\Lambda} u_k(t) e^{ikx}, \tag{4.17}$$

and U_Λ is a finite dimensional projection

$$U_\Lambda(x,t) = \sum_{|k|\leq\Lambda} u_k(t) e^{ikx} = P_\Lambda u(x,t). \tag{4.18}$$

Equation (4.18) is supplemented with the reality condition $u_{-k}(t) = u_k^*(t)$, where u^* denotes complex conjugation. Equation in (4.16) can be recast as a 2Λ-dimensional system of ordinary differential equations

$$\frac{d}{dt}u_k = -\frac{ik}{2} \sum_{\substack{k+p+q=0 \\ |p|,|q|\leq\Lambda}} u_p^* u_q^*, \quad |k| \leq \Lambda. \tag{4.19}$$

Equations in (4.19) conserve energy

$$E = \frac{1}{4\pi} \int_0^{2\pi} P_\Lambda U_\Lambda^2 dx = \sum_{k=1}^{\Lambda} |u_k|^2, \tag{4.20}$$

and Hamiltonian

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

This model was introduced in [61, 62] and statistical properties of this model were further studied in [1].

Since the equation for u_0 is trivial, we can assume $u_0(0) = 0$ without the 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 equi-

librium statistical properties of Fourier coefficients follow a joint Gaussian distribution

$$\pi(u_1, u_2, \dots u_\Lambda) = Ce^{-\beta E}, \tag{4.21}$$

where C is a normalization constant, E is the energy defined in (4.20), and β is the inverse of the temperature determined by the energy of the initial condition. According to the distribution in (4.21) modes u_k achieve equipartition of energy with

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

In [66] it is demonstrated that homogenization can be applied to the TBH model. In particular, equations in (4.19) can be modified by introducing a small parameter ε into the model, and the limit as $\varepsilon \to 0$ of the modified equations can be computed explicitly. Particular approach of introducing the small parameter ε depends on the number of the essential variables in the equation, i.e., the number of modes to be retained in the limit. For a detailed discussion on homogenization for the TBH model see [66]. Here we only consider the case where u_1 is the essential variable. The rest of the variables $k = 2 \dots \Lambda$ 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 following

$$\begin{aligned} \dot{u}_1 &= -\frac{i}{2\varepsilon} \sum_{\substack{p+q+1=0 \\ 2\le|p|,|q|\le\Lambda}} u_p^* u_q^*, \\ \dot{u}_k &= -\frac{ik}{2\varepsilon}\left(u_{k+1}u_1^* + u_{k-1}u_1\right) - \frac{ik}{2\varepsilon^2} \sum_{\substack{k+p+q=0 \\ 2\le|p|,|q|\le\Lambda}} u_p^* u_q^*, \qquad (k \ge 2)\ , \end{aligned} \tag{4.22}$$

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

The limiting behavior of u_1 is given by the SDE

$$da_k = B(a)a_k dt + H(a)a_k dt + \sqrt{2}\sigma(a)dW_k(t), \qquad (k = 1, 2) \tag{4.23}$$

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

$$\begin{cases} B(a) = -(1 - 2N^{-1})\mathcal{E}^{-1}(a)\left(\mathcal{E}^{1/2}(a)I_2|a|^2 + \mathcal{E}^{3/2}(a)I_f\right), \\ \\ H(a) = -N^{-1}\mathcal{E}^{-1/2}(a)|a|^2 I_2 + 2\mathcal{E}^{1/2}(a)I_2 - 3N^{-1}\mathcal{E}^{1/2}(a)I_f, \\ \\ \sigma^2(a) = \mathcal{E}^{1/2}(a)|a|^2 I_2 + \mathcal{E}^{3/2}(a)I_f, \\ \\ \mathcal{E}(a) = N^{-1}(E - |a|^2). \end{cases} \tag{4.24}$$

Details of the derivation are provided in [66]. We point out that the proof of convergence requires assumptions of ergodicity and mixing on the deterministic system in (4.22). Expressions in (4.24) depend on the following parameters; $N = 2\Lambda - 2$ 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 behavior of the fast variables. In particular,

$$\begin{aligned} I_2 &= I\left[\text{Re } u_2, \text{Re } u_2\right] = I\left[\text{Im } u_2, \text{Im } u_2\right], \\ \\ I_f &= I\left[f^{Re}, f^{Re}\right] = I\left[f^{Im}, f^{Im}\right], \end{aligned} \tag{4.25}$$

where $I[\cdot,\cdot]$ is a short-hand notation for the area under the graph of a correlation function (i.e., the *correlation time*)

$$I\left[g,h\right]=\int_0^\infty \langle g(t)h(t+\tau)\rangle_t d\tau, \tag{4.26}$$

where $\langle\cdot\rangle_t$ denotes the temporal average, and

$$\begin{cases} f^{Re}(t) & = \mathrm{Re}\,\Big(-\frac{i}{2}\sum\limits_{\substack{p+q+1=0\\ 2\le|p|,|q|\le\Lambda}} u_p^* u_q^*\Big), \\ \\ f^{Im}(t) & = \mathrm{Im}\,\Big(-\frac{i}{2}\sum\limits_{\substack{p+q+1=0\\ 2\le|p|,|q|\le\Lambda}} u_p^* u_q^*\Big) \end{cases} \tag{4.27}$$

are the terms on the right-hand side of u_1. Parameters I_2 and I_f are estimated from a single microcanonical simulation of the full equations on energy surface $E=2\Lambda$, so that $\langle(\mathrm{Re}\ u_k)^2\rangle_t=\langle(\mathrm{Im}\ u_k)^2\rangle_t=1$. Therefore, 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.

Convergence of the stationary correlation function and the kurtosis in (4.5) for $a_1=\mathrm{Re}\ u_1$ is depicted in Figure 4.3.

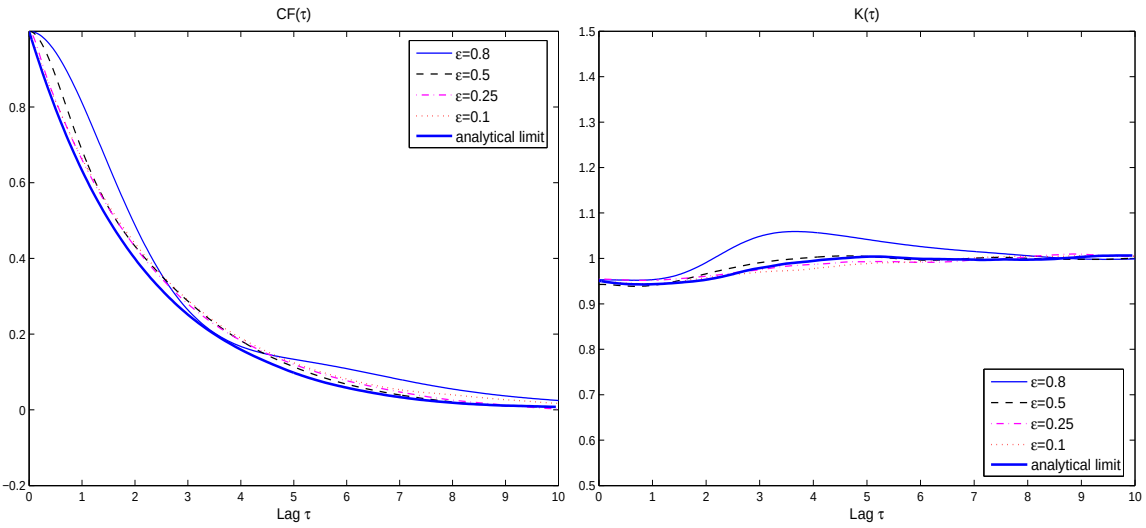

Figure 4.3: Left part - Convergence, as $\varepsilon \to 0$, of correlation function $\langle u_1^{re}(t) u_1^{re}(t+\tau)\rangle_t$ in the modified TBH model (4.22) to the stationary correlation function of a_1 in the effective equation (4.23). Note that the graphs for $\varepsilon = 0.1$ and $\varepsilon = 0.25$ nearly overlap, and thus, are not distinctively visible in the figure. Right part - Convergence, as $\varepsilon \to 0$, of the stationary kurtosis (4.5) for $u_1^{re}(t)$ in (4.22) to stationary kurtosis of a_1 in the effective equation (4.23).

Equations in (4.23) are highly nonlinear, but can be simplified by considering the limit of infinitely many fast variables. In particular, in the limit $N \to \infty$ equations in (4.23) become

$$da_k = b(a)a_k + \sqrt{2}s(a)dW_k, \qquad k = 1, 2, \tag{4.28}$$

where the drift and the diffusion simplify to

$$\begin{aligned} b(a) &= -\left(\sqrt{2\beta}|a|^2 I_2 + (2\beta)^{-1/2} I_f - \sqrt{2/\beta} I_2\right), \\ s^2(a) &= (2\beta)^{-1/2}|a|^2 I_2 + (2\beta)^{-3/2} I_f. \end{aligned} \tag{4.29}$$

It is straightforward to verify that the joint Gaussian density

$$\rho(a) = \frac{1}{2\pi\beta} e^{-\beta|a|^2} \tag{4.30}$$

is a stationary density for the equations in (4.28). We would like to point out that the limits $\varepsilon \to 0$ and $N \to \infty$ do not commute; in particular, values of I_2 and I_f depend on N, but are treated as fixed in the limit of (4.23) as $N \to \infty$. Equation in (4.28) have cubic and linear damping and multiplicative and additive noises. In [66] equations in (4.22) and (4.23) 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 number of fast variables is $N = 2\Lambda - 2 = 38$. Parameters I_2 and I_f become

$$I_2 = 0.14, \;\; I_f = 4.3. \tag{4.31}$$

Although the reduced equations in (4.23) and (4.28) 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 $I_2 \ll 1$ 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. Correlation functions of a_1 in the simulations of the effective SDE in (4.23) and (4.28) and an exponential function $e^{-(2\beta)^{-1/2} I_f t}$ are presented in left part of Figure 4.4. Kurtosis in (4.5) for $a_1 = \text{Re } u_1$ in the simulations of reduced models in (4.23) and (4.28) is presented in the right part of Figure 4.4. Kurtosis in (4.5) is a measure on non-Gaussianity since $Kurt(\tau) = 1$ for Gaussian processes. Behavior of the kurtosis indicates that the non-Gaussian features of both reduced models in (4.23) and (4.28) are extremely weak.

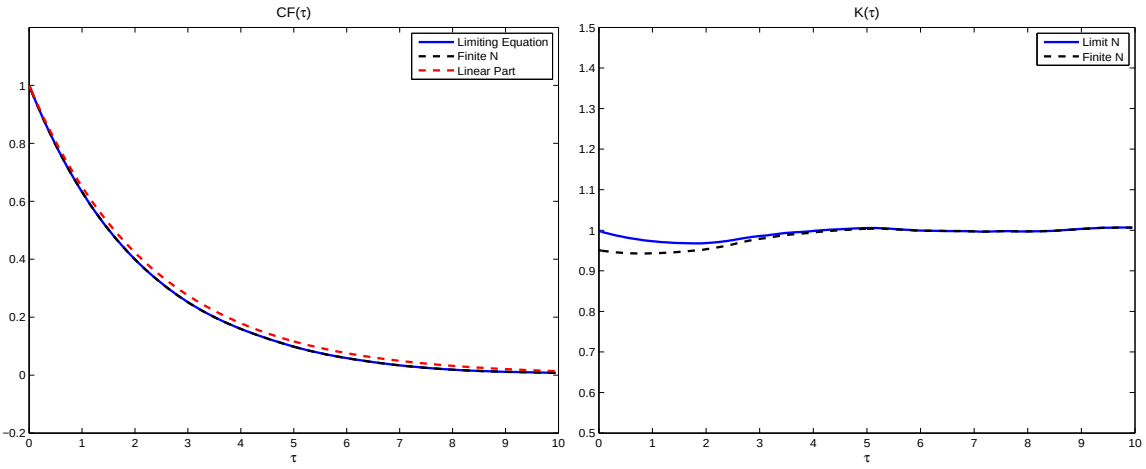

Figure 4.4: Left part - Stationary Correlation function of $a_1 = \text{Re } u_1$ in the effective model (4.23) (black dashed line), limiting equations in (4.28) (solid line), and the exponential function $e^{-(2\beta)^{-1/2} I_f t}$ (red dashed line) which nearly overlaps with the correlation function of the limiting equation. Right part - Stationary kurtosis for $a_1 = \text{Re } u_1$ in the effective model (4.23) (dashed line), and kurtosis for limiting equations in (4.28) (solid line).

Therefore, in the weak energy regime $\beta \gg 1$ system in (4.28) and is well-approximated by the reduced system with $I_2 = 0$. Therefore, in this regime the reduced equation becomes a system of two independent linear Ornstein-Uhlenbeck processes

$$da_k = -\gamma a_k + \sqrt{2} s dW_k, \quad k = 1, 2, \tag{4.32}$$

with $\gamma = (2\beta)^{-1/2} I_f$ and $s^2 = (2\beta)^{-3/2} I_f$. It is easy to verify that linear system in (4.32) has stationary distribution in (4.30).

Therefore, for the purposes of evaluating the performance of estimators in (3.14) we consider that the limiting "true" value of the drift and diffusion coefficients in the OU

approximation of the modified TBH model are

$$\gamma = (2\beta)^{-1/2} I_f, \quad \sigma^2 = 2s^2 = 2(2\beta)^{-3/2} I_f. \tag{4.33}$$

For parameter in (4.31) and $\beta = 50$, coefficients γ and σ^2 become

$$\gamma = 0.43, \quad \sigma^2 = 0.0086. \tag{4.34}$$

In particular, we expect that in the correct subsampling regime estimates for r_0 and r_1 would converge to the values which correspond to the drift and diffusion parameters in (4.33). Nevertheless, we would like to point out that additional assumptions have been made in the derivation of the equation in (4.23). The main simplifying assumption is that the cross-correlation is zero between $a_1 = \text{Re } u_1$ and $a_2 = \text{Im } u_1$. It can be shown numerically that the cross-correlation is extremely weak and very close to zero, but there is no analytical justification for this assumption. This assumption, hence, leads to small discrepancies between the modified model in (4.22) and the reduced model in (4.23). 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 $\varepsilon \to 0$. We demonstrate that our conclusions about subsampling rates and unbiased estimators are still valid in this case.

Sub-sampling Strategy for the Truncated Burgers-Hopf Model

Similar to the triad model considered in the previous section, we test the performance of the estimators in (3.14) for different subsampling strategies by considering several datasets generated by equations in (4.22) with four values of ε. Since the linear effective equations in (4.32) for (a_1, a_2) decouple we consider only parameter estimation in the equation for a_1. Parametric estimation from the data of a_2 follows identical trend. To test the subsam-

pling strategy we consider the dataset $U_n = \text{Re } u_1(n\Delta)$ and compare the behavior of the estimators in (3.14) with the analytical values given by (4.33).

Since the same homogenization procedure is applied to derive the reduce equation in the triad model and the modified TBH model, the time-scale of the fast variables in both models is $O(\varepsilon^2)$. Therefore, behavior of the estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma^2}_\varepsilon$ should be similar to the triad case.

Thus, we consider four subsampling strategies $\Delta = \varepsilon$, $\Delta = \varepsilon^2$, $\Delta = 4\varepsilon^2$, and $\Delta = \varepsilon^3$. Results of parameter estimation for these subsampling strategies are presented in Figure 4.5. Relative errors in the estimator $\hat{\gamma}_\varepsilon$ follow the trend which is identical to the computations from the triad data (c.f., Figures 4.5 and 4.2). In particular, the critical scaling is $\Delta = \varepsilon^2$ and relative errors in the estimation of parameters decay to zero only for the subsampling with $\Delta = \varepsilon^{2\alpha}$, $\alpha \in (0, 1)$.

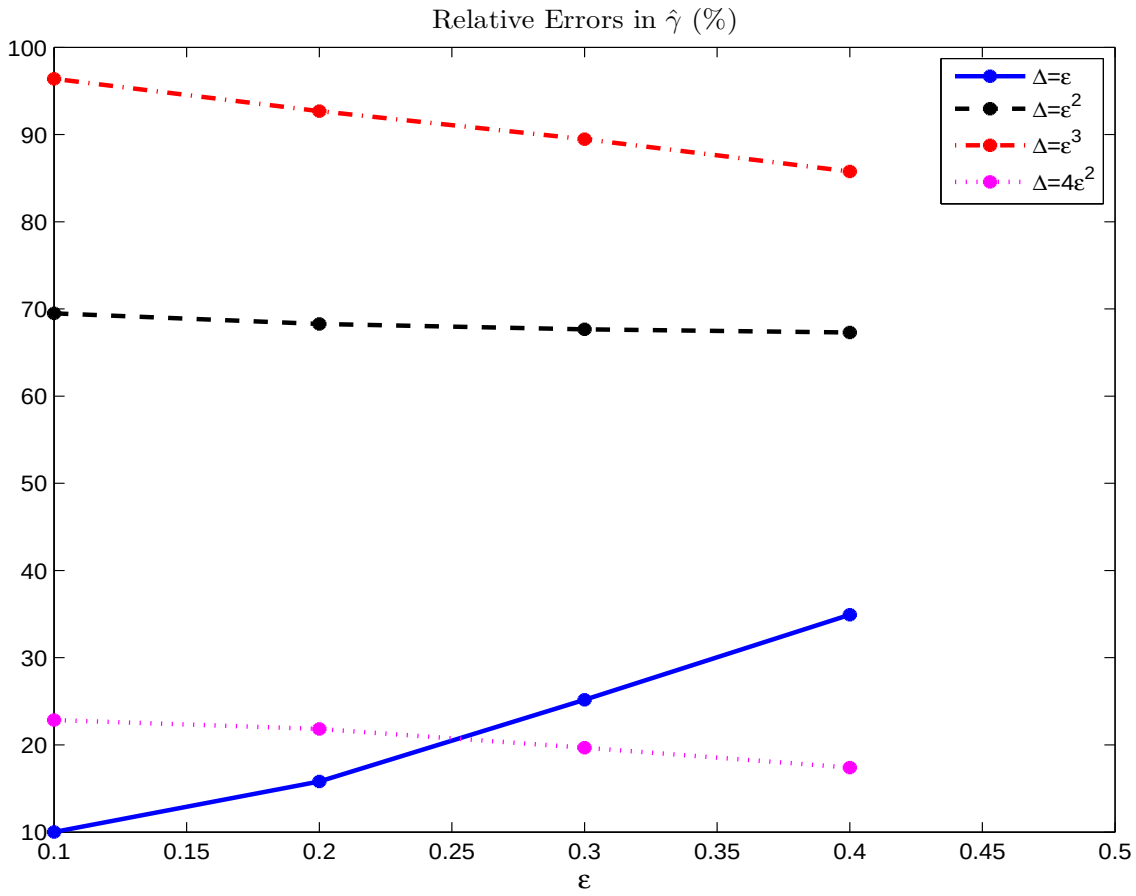

Figure 4.5: Convergence, as $\varepsilon \to 0$, of relative (%) errors in $\hat{\gamma}_\varepsilon$ (3.14) based on observations, from Re $u_1(t)$ in the modified TBH model (4.22), subsampled with four different strategies : solid line - subsampling with $\Delta = \varepsilon$: errors converge to 0, dashed line - subsampling with $\Delta = \varepsilon^2$: errors remain constant, dash-dot line - subsampling with $\Delta = \varepsilon^3$: errors increase to 100%, dotted line - subsampling with $\Delta = 4\varepsilon^2$: errors remain constant.

4.5 Concluding Remarks

Two multiscale dynamic models are presented as prototypical examples for parametric estimation under indirect observability. The corresponding reduced equations (OU process) in the limit of infinite scale-separation ($\varepsilon \to 0$) are known by applying the homogenization method. The objective of consistently estimating the parameters of the reduced equations given the data from the full system is considered. It has been successfully demonstrated that given an increasing number of discretely sampled observations $N(\varepsilon) \gg \Delta/\varepsilon^4$, and selecting $\Delta(\varepsilon) = \varepsilon^{2\alpha}$, for $\alpha \in (0,1)$, ensures L_2-consistency of the (OU) estimators, as $\varepsilon \to 0$.

These results highlight the necessity, as $\varepsilon \to 0$, to subsample the observations from the full multiscale system with a *vanishing but coarse time-step* $\Delta(\varepsilon) \gg \varepsilon^2$ to obtain accurate estimates of the underlying parameters. Precisely, the bias, in particular, in drift estimator $\hat{\gamma}_\varepsilon$, due to the indirect observations is conjectured, with numerical and analytical verification, to be of the form,

$$\hat{\gamma}_\varepsilon - \gamma \sim \text{cte}\, \frac{\gamma \varepsilon^2}{\Delta} \quad \text{for} \quad N \gg \frac{\Delta}{\varepsilon^4}.$$

However, in a practical situation, e.g., when the data has been observed from dynamics with fixed and unknown ε (for instance, when the scale separation is not identifiable), then clearly the bias will be proportional to $\sim (\varepsilon^2/\Delta)$. But no inference regarding the optimality of time-step Δ can be made from the above analytical expression, since the regime $\Delta \gg \varepsilon^2$ is not identifiable. Therefore, with no knowledge of the scale-separation parameter ε the optimal subsampling criterion, verified in this chapter, is not directly indicative of the accuracy of the estimates or the optimal frequency of subsampling. Hence, an alternate approach is required to compute the estimators from a "black box" dynamic model with only indication of the reduced equation. This is addressed next in the chapter.

Chapter 5

Pragmatic Estimation Based on Multiple Subsamples

5.1 Introduction and Motivation

So far we have determined the theoretical limits, as $\varepsilon \to 0$, of the L_2-errors for the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$, based on the observations from approximating processes Y^ε_t, in particular, the SOU process. To obtain meaningful asymptotics the number of discrete observations N and the time step length Δ were assumed to be dependent on the small parameter $\varepsilon > 0$. As $\varepsilon \to 0$, provided $N = N(\varepsilon)$, $\Delta = \Delta(\varepsilon)$ verify certain conditions, the estimators are asymptotically consistent with the best L_2-speeds of convergence. The favorable conditions, or the optimal adaptive subsampling schemes, *are particularly characterised by the small parameter* $\varepsilon > 0$. For instance, for estimation based on SOU process, we recall the following specific optimal characterization (corollary 3.4.3). As $\varepsilon \to 0$, assume that $N(\varepsilon)$ and $\Delta(\varepsilon)$ are given by powers of ε, namely, $N(\varepsilon) = \varepsilon^{-\eta}$, $\Delta(\varepsilon) = \varepsilon^{\alpha}$. Then, as $\varepsilon \to 0$, for any α, η such that $\alpha \in (0, 1)$, $\eta > \alpha$, the OU estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ are asymptotically consistent in L_2 norm. We also show, in chapter 4, that these adequate sub-sampling rates retain their validity in much

wider contexts such as the additive triad system, and the homogenized TBH dynamics.

The above characterization assumes that the exact value of ε is known, which may not be the case in practice. Hence, we consider a different view point under which we assume, that the value of ε is unknown, and moreover, that ε remains fixed. This is relevant to practical situations in which there is only one dataset generated from a numerical simulation of an underlying complex model, for instance, a high dimensional multiscale system, with an unknown fixed parameter ε. Therefore, an important practical task is to develop an approach for analyzing accuracy of the estimators for a particular dataset with a fixed unknown value of ε.

Based on the insights gained from the specific example of the OU/SOU estimation of chapter 3, we will present a straightforward way of estimating the value of the small parameter ε, and in turn identifying the critical scaling $\Delta = \Delta(\varepsilon)$ which minimizes the L_2-error in the estimation.

We illustrate the approach using again the example of the parametric estimation of the OU process under indirect observability framework. The approach involves analysing the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ as functions of the time step Δ, when N is assumed to be large enough, and the dataset of observations from the approximating process is generated with fixed ε. As shown above (chapter 3), asymptotic consistency of the estimators require sub-sampling on a coarse grid, i.e., $\Delta \gg \varepsilon$; therefore, we expect that the errors would increase substantially for $\Delta \to 0$ with ε fixed. This indeed identifies with the condition $\Delta/\varepsilon \to 0$ which is not favorable for consistency, and also under this condition the asymptotic biases (3.19) increase to relative error of 100%.

This is quite often the case when a simplified coarse-grained model, for instance, the OU process, is used to describe the essential features of a large system under investigation, for instance, the additive triad system. Clearly, the mismatch between the *observed* SOU process and the *unobserved* OU process is also of the same nature, as data subsampled with

a high frequency from SOU does not fit the OU process. Such behavior is also common in many applications such as econometric [84, 19] and oceanography [41] when the observed data are contaminated by high frequency observation error. In particular the main difficulty is that the model (OU process) is not compatible with the data at all scales. Therefore, the crucial question is how should the available high frequency data be used for consistent statistical estimation.

In the case of indirect observability estimation based on the SOU process, we will prove that the graph of $\lim_{N\to\infty}(\hat{\gamma}_\varepsilon(N,h)\times h)$ is a straight line as a function of the time step length variable h, for $h \geq \varepsilon$. More specfically, the slope of this straight line is precisely equal to the true γ, and the intercept is used to derive an estimate for the unknown value of ε. This important property provides us with a biased-corrected estimator $\tilde{\gamma}_{slope}$, and an estimate for the unknown parameter ε.

We will show that this is a *useful universal approach* to address the issue of finding the critical subsampling required for precise estimates, especially, when the small parameter ε is fixed and unknown.

5.2 Parametric Estimation under Indirect Observability with a Fixed Unknown ε

We recall the definition and a few properties of SOU process Y_t^ε from chapter 3. The process Y_t^ε is obtained by averaging the OU process X_t over a sliding window of fixed length $\varepsilon > 0$, so that

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

The associated covariance function $K^{\varepsilon}(h)$ for $h \geq 0$, given by (3.4), takes the following form,

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

From proposition 3.3.2, the estimators $\hat{\gamma}_{\varepsilon}$ and $\hat{\sigma}^2_{\varepsilon}$ (3.14) based on discrete observations subsampled from SOU process verify the following convergence in L_2, for fixed values of ε and Δ (*fixed rate subsampling*), and as $N \to \infty$,

$$\hat{\gamma}_{\varepsilon} \to G = G(\varepsilon, \Delta), \qquad \hat{\sigma}^2_{\varepsilon} \to S = S(\varepsilon, \Delta), \tag{5.1}$$

where

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

and where the covariances $K^{\varepsilon}(0)$ and $K^{\varepsilon}(\Delta)$ are given by (3.4).

Now, for fixed small value of ε, we define a function $\Gamma(h) = G(\varepsilon, h) \times h$, for $h \geq 0$, such that,

$$\Gamma(h) = \log\left(K^{\varepsilon}(0)/K^{\varepsilon}(h)\right). \tag{5.2}$$

Then, $\Gamma(h)$ is differentiable with respect to h, and its derivative is given by,

$$\frac{d\Gamma(h)}{d\, h} = \begin{cases} \gamma + \dfrac{2\gamma e^{\gamma h}\left(1 - \gamma(\varepsilon - h) - e^{\gamma(h-\varepsilon)}\right)}{2\gamma(\varepsilon - h)e^{\gamma h} + e^{-\gamma\varepsilon}(e^{2\gamma h} + 1) - 2}, & 0 \leq h < \varepsilon, \\ \gamma, & h \geq \varepsilon. \end{cases} \tag{5.3}$$

Similarly, we define $\Sigma(h) = S(\varepsilon, h) \times h = 2\, K^{\varepsilon}(0)\Gamma(h)$, to obtain,

$$\frac{d\Sigma(h)}{d\, h} = \begin{cases} \sigma^2 C_0 \left(1 + \dfrac{2e^{\gamma h}\left(1 - \gamma(\varepsilon - h) - e^{\gamma(h-\varepsilon)}\right)}{2\gamma(\varepsilon - h)e^{\gamma h} + e^{-\gamma\varepsilon}(e^{2\gamma h} + 1) - 2}\right), & 0 \le h < \varepsilon, \\ \\ \sigma^2 C_0, & h \ge \varepsilon, \end{cases} \tag{5.4}$$

where $C_0 = 2(\gamma\varepsilon + e^{-\gamma\varepsilon} - 1)/(\gamma^2\varepsilon^2)$. The derivative, in particular, of $\Gamma(h)$ remains constant with respect to h for $h \ge \varepsilon$. Therefore, for $h \ge \varepsilon$, using (5.3) we obtain

$$\Gamma(h) = \Gamma(\varepsilon) + \gamma(h - \varepsilon) = \gamma\, h + C, \tag{5.5}$$

where the intercept $C = \Gamma(\varepsilon) - \gamma\varepsilon$.

Result 5.2.1 *Given N observations from SOU process subsampled with time step h, we consider the erstwhile OU estimator $\hat{\gamma}_\varepsilon = \hat{\gamma}_\varepsilon(N, h)$, given by* (3.14) *with $\Delta = h$. Then,* (5.1), (5.2) *and* (5.5) *prove that the estimator $(\hat{\gamma}_\varepsilon \times h)$ converges in L_2 to $(\gamma\ h + C)$ as $N \to \infty$, for each fixed $h \ge \varepsilon$. The intercept $C = -(\gamma\varepsilon/3)(1 + O(\varepsilon))$, as $\varepsilon \to 0$. Hence, for the fixed small value of ε, we have, $C \approx -\gamma\varepsilon/3$.*

Based on the result 5.2.1 we develop a numerical methodology to estimate OU parameters γ, σ^2 given observations from an approximating process with a fixed unknown value of ε. We verify the approach in the numerical computations presented below for the *smoothed Ornstein-Uhlenbeck process*. The approach proves fruitful in the estimation of parameters based on the observations from the *Additive Triad system*, and the TBH model as well.

5.2.1 Bias-corrected Estimators and Estimate for Unknown ε

The observations made in previous section are used to develop the following methodology to compute estimators for γ, σ^2. For brevity, we will consider that estimation is based on dis-

crete observations from the SOU process. However, we conjecture that the methodology is replicable in more general situations. As mentioned above, we will show numerical verifications in the two examples, namely, the homogenized multiscale systems, for the applicability of our approach.

Given N observations subsampled from the approximating SOU process $U_n = Y^{\varepsilon}_{n\,h}$ at uniform time step $h > 0$. Assume that the value of ε is fixed, and is to be estimated. Consider the OU estimators $\hat{\gamma}_\varepsilon = \hat{\gamma}_\varepsilon(N,h)$ and $\hat{\sigma}^2_\varepsilon = \hat{\sigma}^2_\varepsilon(N,h)$, given by (3.14) for $\Delta = h$. Since $\hat{\sigma}^2_\varepsilon = 2K^\varepsilon(0)\hat{\gamma}_\varepsilon$; therefore, we focus only on the estimation of the drift coefficient γ.

Define $\hat{\Gamma}(h) = \hat{\Gamma}(h;N,\varepsilon) = \hat{\gamma}_\varepsilon \times h$, then, by (5.1) (5.2), $\hat{\Gamma}(h) \to \Gamma(h)$ as $N \to \infty$ in L_2, for each h. Using result 5.2.1, it is inferred that a portion of the curve $\lim_{N\to\infty} \hat{\Gamma}(h)$ is of constant slope, namely, for $h \geq \varepsilon$. Since ε is assumed to be unknown, and we only have finitely many but large N, the constant slope portion of the curve is not known apriori. Therefore, for large values of N, the objective is to fit a straight line to a part of the curve $\hat{\Gamma}(h)$ sampled at discrete h values which minimizes the least square errors, i.e., assuming we have values of $\hat{\Gamma}(h)$ at finitely many time steps $h = h_1 < h_2 < \ldots < h_M$, then,

$$\hat{\Gamma}(h_j) = \hat{\gamma}_\varepsilon(N,h_j)\, h_j = a\, h_j + b + e_j, \tag{5.6}$$

where $e_j = e(N,h_j)$ are the residuals and a, b are to be estimated. Then, ordinary least square estimation provides us with estimates $\hat{a}$, $\hat{b}$ for a and b. To select the "optimal" portion of the discrete curve $\{\hat{\Gamma}(h_1),\ldots,\hat{\Gamma}(h_M)\}$, we select some number m from the M time step values, say $\{h_{i+1} < h_{i+2} < \ldots < h_{i+m}\}$, such that $i = 0,1,\ldots,M-m$. Then, estimate the model (5.6) with $(M-m)$ datasets $\{h_{i+j}, \hat{\Gamma}(h_{i+j})\}_{j=1,m}$. The value of estimates $\hat{a}$, $\hat{b}$ that minimize the residual sum of squares will be thus selected. Let us say we denote by $\hat{a}$, $\hat{b}$ the final estimates that are selected using the approach described here. Then in view of the

result 5.2.1, we define the following estimator for the parameter γ,

$$\hat{\gamma}_{slope} = \hat{a}, \tag{5.7}$$

and an estimate for the unknown ε by, $\hat{\varepsilon} = -3\hat{b}/\hat{\gamma}_{slope}$. At this step, one may initialize the step of identifying the parameters a, b in (5.6), by considering only those h_j which are greater than $\hat{\varepsilon}$. Then, we would define $\hat{\sigma}^2_{slope} = 2\hat{K}^{\varepsilon}(0)\hat{\gamma}_{slope}$.

Complementing the above approach, we derive another methodology in the same spirit as above, and define an estimator for drift parameter γ. Let $h_1, h_2 > 0$ be two distinct values of the sampling time step, then consider the estimators $\hat{\gamma}_{\varepsilon}(N, h_1)$ and $\hat{\gamma}_{\varepsilon}(N, h_2)$, given by (3.14) for $\Delta = h_1, h_2$, and the associated values of $\hat{\Gamma}(h_1)$ and $\hat{\Gamma}(h_2)$. Then, define a *two-scale* estimator $\tilde{\gamma}_{\varepsilon} = \tilde{\gamma}_{\varepsilon}(h_1, h_2, N)$ as,

$$\tilde{\gamma}_{\varepsilon} = \frac{\hat{\Gamma}(h_1) - \hat{\Gamma}(h_2)}{h_1 - h_2} \to \frac{\Gamma(h_1) - \Gamma(h_2)}{h_1 - h_2} = \gamma, \quad \text{provided } h_1,\ h_2 \geq \varepsilon, \tag{5.8}$$

here the convergence (by proposition 3.3.2) is in the sense of L_2 under the fixed rate sub-sampling, i.e., $N \to \infty$, such that time steps h_1, h_2 remain fixed. For large number of observations N, and for h_1 close to h_2, clearly, estimator $\tilde{\gamma}_{\varepsilon}$ is the finite difference approximation to the derivative of $\hat{\Gamma}(h)$ at h_1. Therefore, given values of $\hat{\Gamma}(h)$ at finitely many time steps $h = h_1 < h_2 < \ldots < h_M$, compute the *two-scale* estimators $\tilde{\gamma}_{\varepsilon}(h_i, h_{i+1})$ (dependence on N is supressed) for every consecutive pair of time steps $h_i < h_{i+1}$, then, for values of h_i larger than the unknown ε, the estimators $\tilde{\gamma}_{\varepsilon}(h_i, h_{i+1})$ converge in L_2 to γ as $N \to \infty$. Also, for values of $h_i \ll \varepsilon$, $\tilde{\gamma}_{\varepsilon}$ will be close to zero, since, the derivative $\Gamma(h)$ for $h \ll \varepsilon$ is close zero. Therefore, the maximal h_i such that the estimators $\tilde{\gamma}_{\varepsilon}$ stabilize around a constant value for all $h \geq h_i$, is a viable estimate for the unknown value of ε. This notion will be exploited in the numerical simulations presented below to analyze the estimators and compute precise estimates for γ, and hence, σ^2.

Note that if the value of the small parameter ε is known, then the bias-corrected estimators, for instance, $\hat{\gamma}_{slope}$ based on multiple time steps $h_i > \varepsilon$ will be our best estimator.

5.3 Numerical Results for the SOU Process

The conventional technique to deal with situations when one fits a coarse-grained model to high frequency data is to subsample the data with a larger time step, i.e., sample on a coarse grid. This reduces the bias in the estimation. In particular, recall the expressions for the biases in $\hat{\gamma}_\varepsilon$ given in (3.20), (3.21), namely,

$$Bias_\gamma = -(1/\Delta)\ln\left(\frac{e^{-\gamma\varepsilon}+e^{\gamma\varepsilon}-2}{2\left(e^{-\gamma\varepsilon}-1+\gamma\varepsilon\right)}\right) = -\frac{\gamma\varepsilon}{3\Delta}\left(1+\frac{\gamma\varepsilon}{6}+O(\varepsilon^2)\right), \qquad \Delta \geq \varepsilon,$$

$$Bias_\gamma = -(1/\Delta)\ln\left(\frac{2\gamma(\varepsilon-\Delta)e^{\gamma\Delta}+e^{-\gamma\varepsilon}(e^{2\gamma\Delta}+1)-2}{2\left(e^{-\gamma\varepsilon}-1+\gamma\varepsilon\right)}\right), \qquad \Delta < \varepsilon.$$

From these expressions, we infer that as Δ increases then the bias does reduce, but it is not clear which value of Δ is optimal. Neither can we infer the value of the small parameter ε from the graph of the $Bias_\gamma$ plotted against the time step Δ.

Consider the figure 5.1 that has been generated by taking the following values,

$$\varepsilon = 0.01, \quad \gamma = 3.2625, \quad \sigma = 6.7500.$$

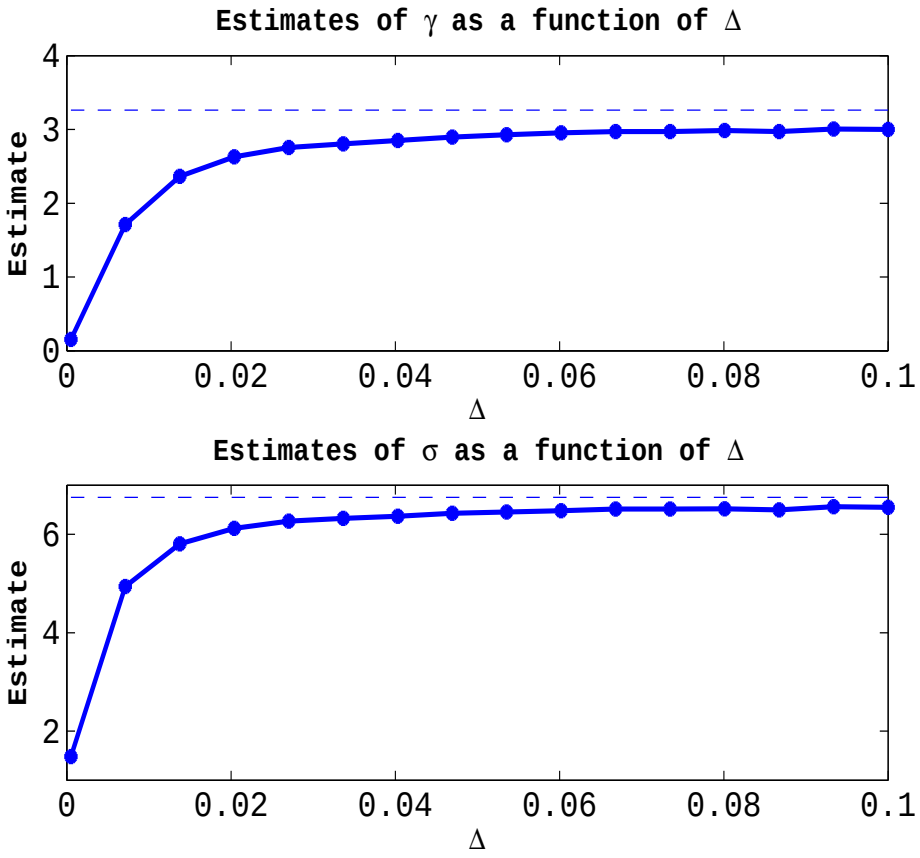

Figure 5.1: Estimation based on SOU dataset with $\varepsilon = 0.01$. Top Plot: Solid line - $\hat{\gamma}_\varepsilon$ given by (3.14) as a function of Δ, dashed line - true value of $\gamma = 3.2625$. Bottom Plot: Solid line - $\hat{\sigma}_\varepsilon$ given by (3.14) as a function of Δ, dashed line - true value of $\sigma = 6.75$.

The figure plots the estimator $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ as a function of Δ. The estimators increase to the corresponding true values of the parameters, but the optimal subsampling time step value is not evident from Figure 5.1.

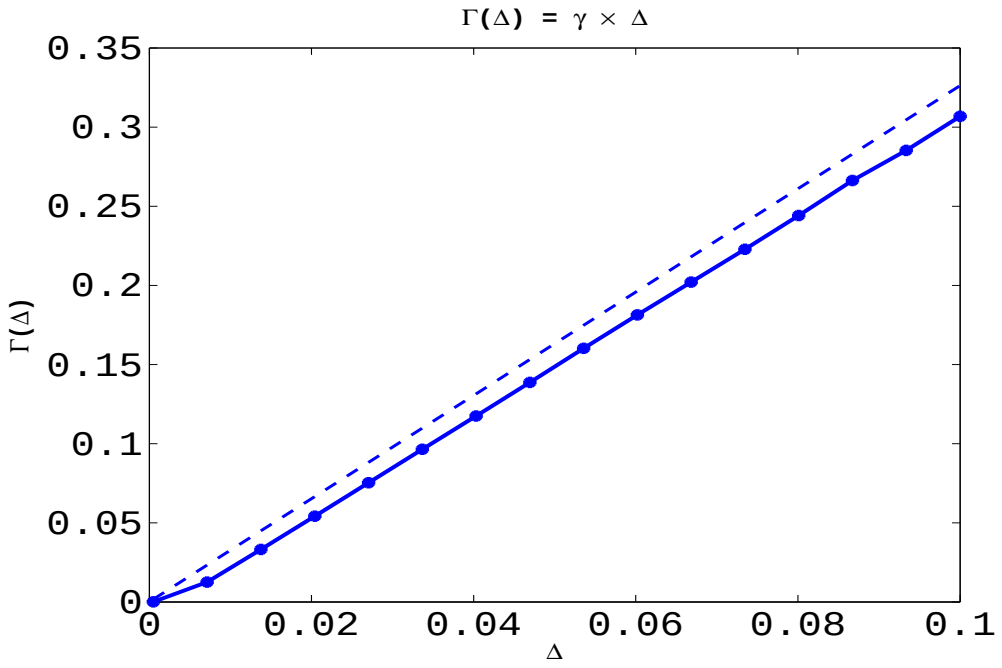

Figure 5.2: Estimation based on SOU dataset with $\varepsilon = 0.01$. Solid line - $\hat{\Gamma}(\Delta)$ given by (5.6) as function of Δ, dashed line - straight line through origin with slope $\gamma = 3.2625$.

Figure 5.2 plots $\hat{\Gamma}(\Delta)$ for different values of Δ. As shown above, for large N the curve $\hat{\Gamma}(\Delta)$ is a straight line for $\Delta \geq \varepsilon$. This is clearly evident from Figure 5.2. Also, we obtain that the estimator $\hat{\gamma}_{slope} = 3.1831$, when we fit (5.6) to the appropriate data. The estimate for the small parameter is $\hat{\varepsilon} = 0.0101$.

The *two-scale* estimators $\tilde{\gamma}_\varepsilon$ (5.8) are plotted in Figure 5.3 against different values of Δ by considering forward finite differences. As shown above, for large value of N, the estimator $\tilde{\gamma}_\varepsilon$ should stabilize around a constant value for $\Delta > \varepsilon = 0.01$, which can be easily deduced from the Figure 5.3.

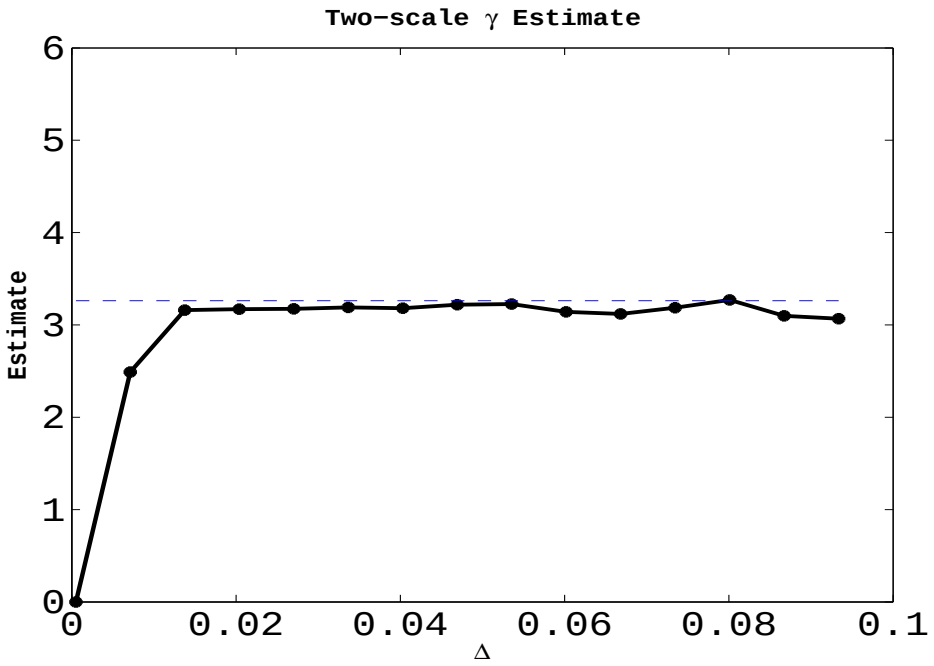

Figure 5.3: Two-scale estimator (5.8) based on SOU dataset with $\varepsilon = 0.01$. Solid line - $\tilde{\gamma}_\varepsilon$ as a function of Δ, dashed line - true value $\gamma = 3.2625$.

5.4 Numerical Results for the Additive Triad System

In chapter 4 we utilized datasets generated from numerical simulations of the multiscale additive triad system in (4.1) with known values of ε to test the behavior of the bias and to elucidate the correct sub-sampling strategy which guarantees that the bias converges to zero. We verified that the critical optimal subsampling regime is characterized by $\Delta \gg \varepsilon^2$, and that in the optimal subsampling regime the errors in the estimation (4.7) are proportional to (ε^2/Δ) (Figure 4.2). Therefore, our conjecture that ε^2 plays the role of ε in the SOU process gained evidence.

As discussed above, in practice, there is only one dataset which is generated from a

numerical simulation of a complex model without an explicit parameter ε. Thus value of ε is, typically, not known, nevertheless it is desirable to model the dataset with a particular reduced stochastic model. Therefore, an important practical task is to develop an approach for analyzing estimators' accuracy for a particular dataset with a fixed finite value of ε.

Now we apply the approach developed in section 5.2.1, to estimation of OU parameters under indirect observability with data subsampled from the triad system.

In particular, we conjecture that given data from the triad system (4.1) for a fixed value of ε, if we consider the drift estimator (3.14) as functions of time step $h > 0$, i.e. $\hat{\gamma}_\varepsilon = \hat{\gamma}_\varepsilon(h)$, then the following relationship holds, as $N \to \infty$

$$\hat{\Gamma}(h) = \hat{\gamma}_\varepsilon(h)h \to \gamma h + C_1(\varepsilon), \qquad \text{for} \quad h > \varepsilon^2, \tag{5.9}$$

where C_1 is a constant depending on the unknown fixed value of ε. Similar asymptotic relation will hold for the diffusion estimator $\hat{\sigma}^2_\varepsilon$.

We utilize the triad data with $\varepsilon = 0.3$ to elucidate the behavior of the estimators for a small, but finite, value of ε. In particular, behavior of $\hat{\gamma}_\varepsilon(\Delta)$ and $\hat{\Gamma}(\Delta) = (\hat{\gamma}(\Delta) \times \Delta)$ for varying Δ is depicted in Figure 5.4. Numerical results depicted in left part of Figure 5.4 show a typical hyperbolic profile consistent with the expression in (4.7). The critical scaling in this case is $\varepsilon^2 = 0.09$, but we would like to point out that the estimator $\hat{\gamma}_\varepsilon$ is biased, with an order $O(\varepsilon^2/\Delta)$, for any value of Δ, including $\Delta > 0.09$, i.e., in the optimal subsampling regime. Moreover, considering the graph $(\Delta, \hat{\gamma}_\varepsilon)$ for a given dataset with fixed ε (left part of Figure 5.5) does not, by itself, provide any significant information about the value of ε.

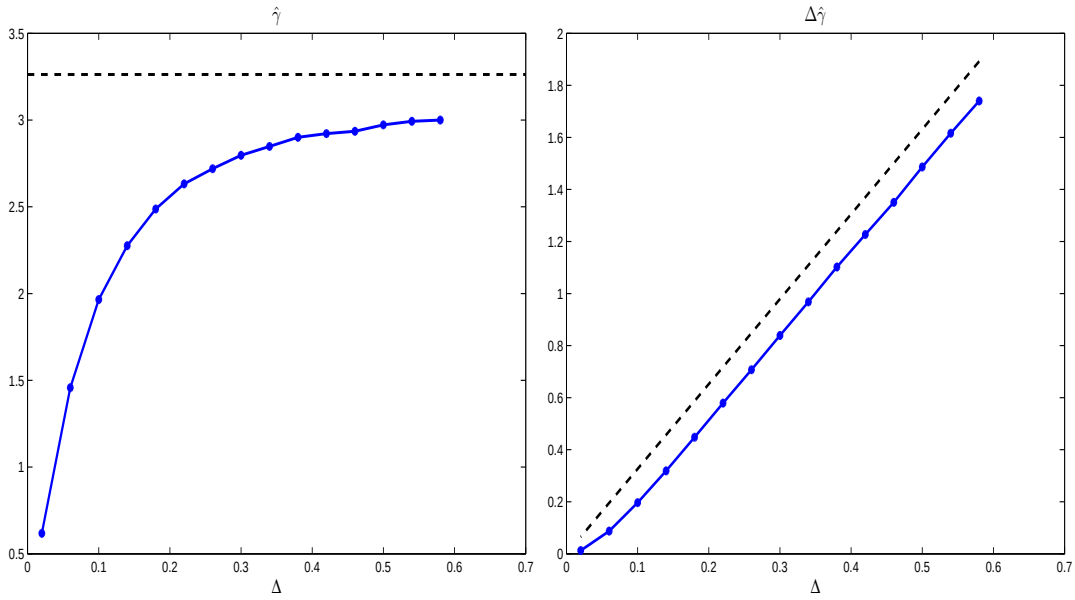

Figure 5.4: Left part - estimator $\hat{\gamma}(\Delta)$ for different values of Δ computed from the TRIAD dataset with $\varepsilon = 0.3$. Solid line - Estimator $\hat{\gamma}_\varepsilon(\Delta)$ given by (3.14), dashed line - analytical asymptotic true value $\gamma = 3.2625$ computed from (4.4). Right part - $\hat{\Gamma}(\Delta) = (\hat{\gamma}_\varepsilon(\Delta) \times \Delta)$ as a function of Δ estimated from the TRIAD dataset with $\varepsilon = 0.3$. Solid line - $\hat{\Gamma}(\Delta)$, dashed line - straight line with the slope $\gamma = 3.2625$ given by the analytical formula (4.4).

As conjectured, $\hat{\Gamma}(\Delta)$ becomes approximately a straight line for $\Delta > \varepsilon^2$, such that $\varepsilon^2 = 0.09$ in this example. The region where $\Gamma(\Delta)$ does not follow a linear relationship (5.9) is clearly identifiable from the Figure 5.4. Therefore, this simple diagnostics allows to estimate the critical value of the small-scale parameter ε^2. Now we consider the first approach of fitting a straight line (5.9) to $\hat{\Gamma}(\Delta)$ in the region $\Delta > \varepsilon^2$ for the triad, where the value of ε^2 is clear from the Figure 5.4. The slope of this straight line yields the estimator $\hat{\gamma}_{slope}$ for the drift parameter γ. To use the linear regression approach we neglect first few data points for small values of Δ (in particular, we neglect data for $\Delta = 0.02,\ 0.06$) and compute the

estimator $\hat{\gamma}_{slope}$ given by the coefficient of the linear regression,

$$\hat{\gamma}_{slope} = 3.23.$$

The relative error for the regression estimator is approximately 1%. This is a remarkable improvement compared to the standard estimator $\hat{\gamma}_\varepsilon$. Note that the regression estimator is much more accurate compared with the estimator $\hat{\gamma}_\varepsilon$ even for large values of Δ, for instance, for $\Delta = 0.58 > \varepsilon^2$ the corresponding value for $\hat{\gamma}_\varepsilon \approx 3$.

The second complementary estimator that we consider is the *two-scale* estimator $\tilde{\gamma}_\varepsilon$ given by (5.8), which is the numerical approximation of the slope (synonymously, the derivative) of $\hat{\Gamma}(\Delta) = (\hat{\gamma}_\varepsilon(\Delta) \times \Delta)$. For this reason we denote it, by a slight abuse of notation, as

$$\tilde{\gamma}_\varepsilon = \frac{d}{d\Delta}\left(\hat{\Gamma}(\Delta)\right). \tag{5.10}$$

According to the linear relationship (5.9) the estimators $\tilde{\gamma}_\varepsilon$ and, with similar definition, $\tilde{\sigma}^2_\varepsilon$ are unbiased estimators for $\Delta > \varepsilon^2$. We expect $\tilde{\gamma}_\varepsilon \approx \gamma$ and $\tilde{\sigma}^2_\varepsilon \approx \sigma^2$ for values of Δ in the consistent sub-sampling regime, i.e. where the discrete data from the triad well-approximates a discrete sample from the effective Ornstein-Uhlenbeck process. We compute the two-scale estimator $\tilde{\gamma}_\varepsilon$ at each discrete value of Δ by taking the central difference numerical approximation for the derivative in (5.10).

Numerical estimates for (5.10) are depicted in Figure (5.5). Results presented in Figure 5.5 demonstrate that numerical approximation for $\tilde{\gamma}_\varepsilon$ in (5.10) stabilizes in the correct regime, i.e. $\Delta \geq 0.2$ while for this dataset the critical scaling is $\varepsilon^2 = 0.09$.

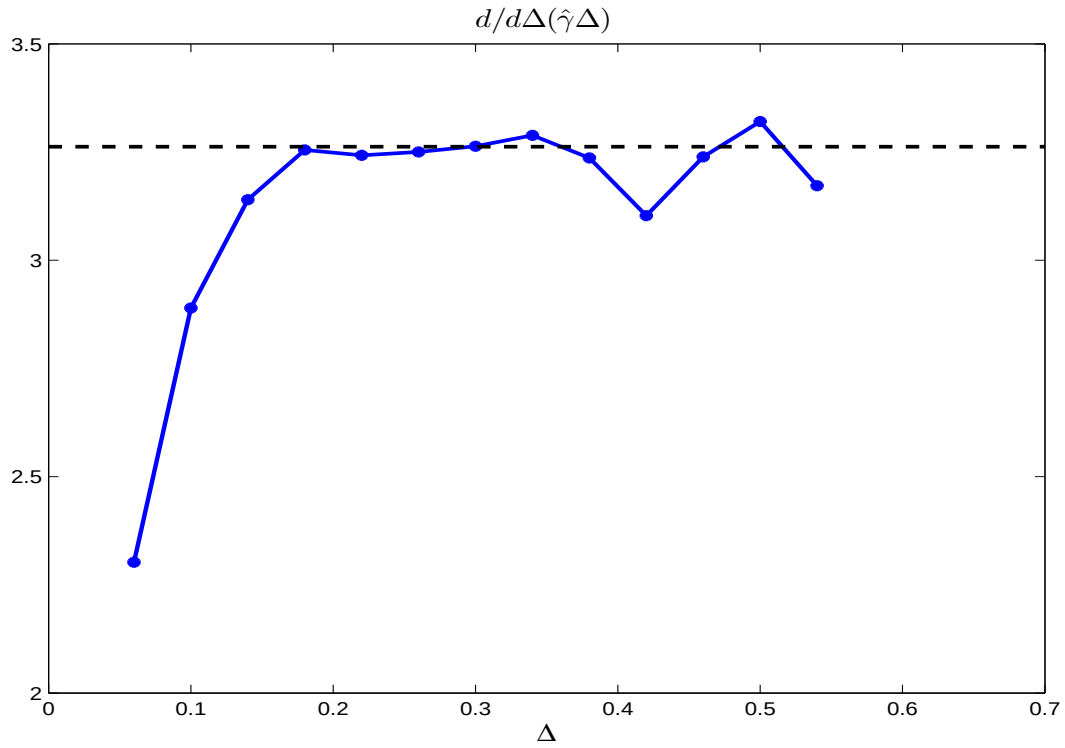

Figure 5.5: Estimator $\tilde{\gamma}_\varepsilon$ in (5.10) as a function of Δ, computed from the TRIAD dataset with $\varepsilon = 0.3$. Solid line - $\tilde{\gamma}_\varepsilon$ in (5.10), dash-dot line - analytical asymptotic true value $\gamma = 3.2625$ computed from (4.4).

The central finite-difference approximation for the $\tilde{\gamma}_\varepsilon$ in (5.10) yield accurate estimation of γ up to a very high precision for intermediate values of Δ. Higher numerical errors for larger values of $\Delta \approx 0.45 \ldots 0.6$ are due to the poor finite-difference approximation of the derivative. Therefore, expression in (5.10) provides a practical expression for determining a precise estimate for the parameter under investigation, and for estimating the value of the critical subsampling time step $\Delta > \varepsilon^2$.

5.5 Numerical Results for the Truncated Burgers-Hopf Equations

Similar to the discussion in previous section, we also analyze dependence of the estimator $\hat{\gamma}_\varepsilon$ on Δ for a particular dataset with a fixed unknown value of ε. In particular, $\hat{\gamma}_\varepsilon(\Delta)$ and $\hat{\Gamma}(\Delta) = \Delta \times \hat{\gamma}(\Delta)$, as functions of Δ, computed from the dataset subsampled from the simulations of the modified TBH with $\varepsilon = 0.3$ are depicted in Figure 5.6. Similar to the triad case presented in the previous section, considering the graph of $\hat{\gamma}_\varepsilon(\Delta)$ as a function of Δ provides only limited information about the correct sub-sampling strategy. Estimator $\hat{\gamma}(\Delta)$ as a function of the sub-sampling step, Δ, is presented in the left part of Figure 5.6. Considering $\hat{\gamma}(\Delta)$ vs Δ illustrates that the estimator $\hat{\gamma}_\varepsilon$ is sensitive to changes in Δ even for very large $\Delta \approx 0.5$. On the other hand, plot of $\hat{\Gamma}(\Delta)$ vs Δ presented in the right part of Figure 5.6 provides a much better estimate for the value of the small parameter in the problem. The curve $\hat{\Gamma}(\Delta)$ 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. Considering $\Delta\hat{\gamma}(\Delta)$ as a function of Δ also allows to construct an unbiased estimator as a linear regression fit. In particular, plot $\Delta\hat{\gamma}(\Delta)$ vs Δ suggests that four points for small values of Δ should be neglected in considering a regression fit. After neglecting the first four poins the regression coefficient for the straight line in the right part of Figure 5.6, for the fixed value of $\varepsilon = 0.3$, becomes

$$\hat{\gamma}_{slope} \approx 0.45.$$

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

We also test the performance of the two-scale estimators $\tilde{\gamma}_\varepsilon$ introduced in (5.10) and (5.8). In this case behavior of the estimator $\tilde{\gamma}_\varepsilon$ is slightly different compared with the triad case. The estimator $\tilde{\gamma}_\varepsilon$ does not stabilize to a constant in the range $\Delta \gg \varepsilon^2$; instead it varies slightly and even decays for $\Delta > 0.3$. Nevertheless, it is beneficial to consider estimators $\hat{\gamma}_\varepsilon$, $\hat{\gamma}_{slope}$, and $\tilde{\gamma}_\varepsilon$ simultaneously. Together with $\hat{\Gamma}(\Delta)$ they provide crucial insight into the fast time-scale in the problem and estimating precise value of the parameters.

As a second test case, we also illustrate the behavior of $\hat{\gamma}_\varepsilon(\Delta)$ and $\hat{\sigma}^2_\varepsilon(\Delta)$ as functions of Δ, for fixed $\varepsilon = 0.1$. In particular, $\hat{\Gamma}(\Delta) = (\Delta \times \hat{\gamma}(\Delta))$ and $\hat{\Sigma}(\Delta) = (\Delta \times \hat{\sigma}^2(\Delta))$ are plotted in Figure (5.8). There is a noticable change in the curvature of both lines at $\Delta \approx 0.01$. This value of Δ corresponds to the critical scaling $\Delta = \varepsilon^2$. Therefore, the fast time-scale is clearly identifiable from both graphs. Moreover, linear regression provides the following values for the slopes of two lines

$$\hat{\gamma}_{slope}(\varepsilon = 0.1) \approx 0.415, \quad \hat{\sigma}^2_{slope}(\varepsilon = 0.1) \approx 0.009, \tag{5.11}$$

such that the true values are given by (4.34), namely, $\gamma = 0.43$, $\sigma^2 = 0.0086$. We plot the three estimators (biased estimator in (3.14), linear regression estimator, and the two-scale estimator in (5.10)) computed from the TBH dataset with $\varepsilon = 0.1$ on the same graph in Figure (5.9). Figure 5.9 demonstrates that the consistent sub-sampling regime is clearly identifiable in Figure 5.9 from the behavior of $\tilde{\gamma}$ and $\tilde{\sigma}^2$. Moreover, numerical values of parameters can be estimated within 5% relative errors by comparing the behavior of the biased estimators in (3.14), regression fitting, and two-scale estimators in (5.10).

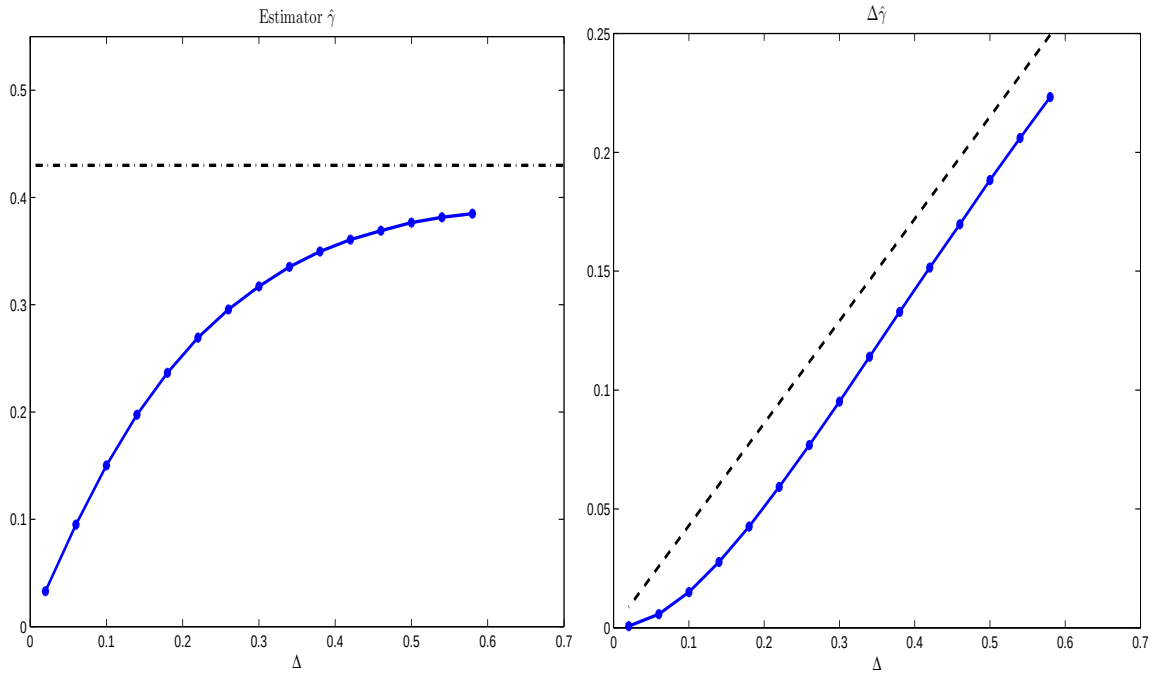

Figure 5.6: Left part - estimator $\hat{\gamma}_\varepsilon$ vs Δ computed from the TBH dataset with $\varepsilon = 0.3$. Solid line - $\hat{\gamma}_\varepsilon(\Delta)$ in (3.14), dash-dot line - analytical asymptotic value in (4.33). Right part - $\hat{\Gamma}(\Delta) = (\Delta \times \hat{\gamma}_\varepsilon)$ in (5.6) computed from the TBH dataset with $\varepsilon = 0.3$, dashed line - $(\Delta \times \gamma)$ where γ is the analytical asymptotic value in (4.33).

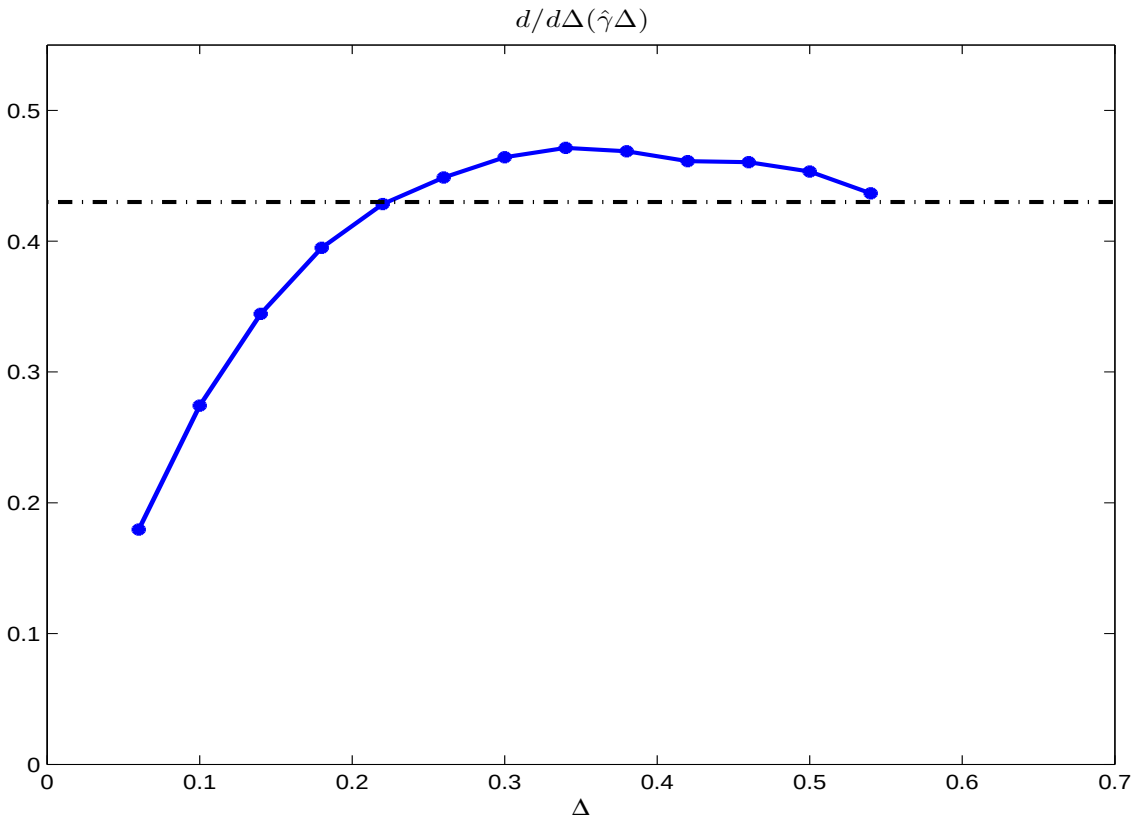

Figure 5.7: Estimator $\tilde{\gamma}_\varepsilon$ in (5.10) computed from the TBH dataset with $\varepsilon = 0.3$, dash-dot line - analytical asymptotic value in (4.33).

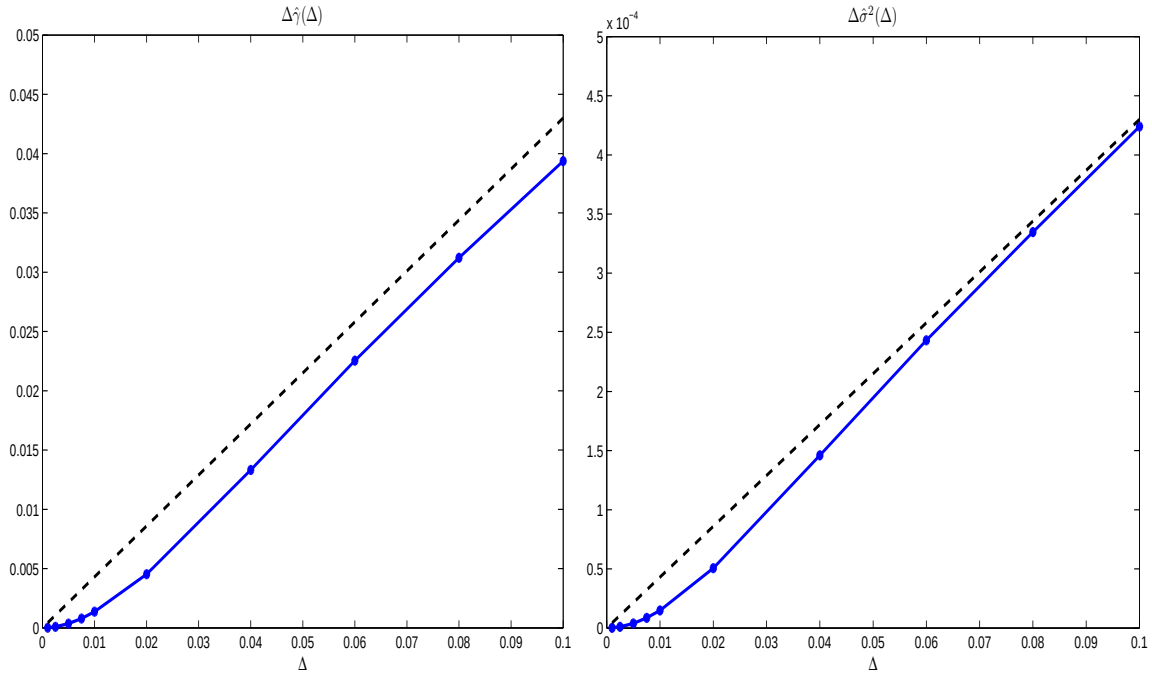

Figure 5.8: Behavior of $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}^2_\varepsilon$ as functions of Δ for data computed from the TBH dataset with $\varepsilon = 0.1$. Left part: solid line - $(\Delta \times \hat{\gamma}_\varepsilon(\Delta))$ vs Δ, dashed line - straight line with the slope γ in (4.33). Right part: solid line - $(\Delta \times \hat{\sigma}^2_\varepsilon(\Delta))$ vs Δ, dashed line - straight line with the slope σ^2 in (4.33). Values for $\Delta > 0.1$ are not shown to emphasize the behavior near $\Delta = \varepsilon^2 = 0.01$.

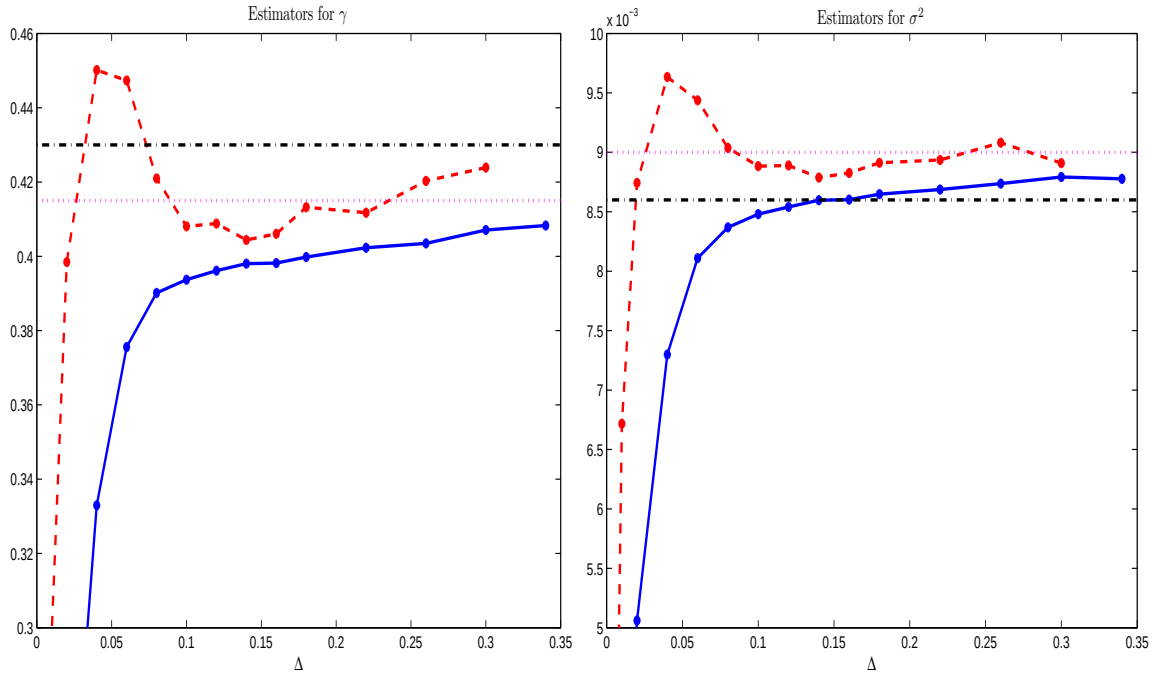

Figure 5.9: Estimators as functions of Δ computed from the TBH dataset with $\varepsilon = 0.1$. Left part: solid line - $\hat{\gamma}_\varepsilon(\Delta)$ in (3.14), dashed line - estimator $\tilde{\gamma}_\varepsilon$ in (5.10), dotted line - $\hat{\gamma}_{slope}$ in (5.11), dash-dot line - analytical asymptotic value in (4.33). Right part: solid line - $\hat{\sigma}^2_\varepsilon(\Delta)$ in (2.25), dashed line - $\tilde{\sigma^2}_\varepsilon$ in (5.10), dotted line - $\hat{\sigma}^2_{slope}$ in (5.11), dash-dot line - analytical asymptotic value in (4.33).

5.6 Concluding Remarks

A methodology to accurately estimate parameters from a given dataset sampled from an approximating process, for instance, the SOU process, or the slow variables in a multiscale system, is presented. The observed dataset is assumed to be sampled from a trajectory generated with a fixed, and unknown value of small parameter ε.

The optimal subsampling regimes identified by conditions such as $\Delta \gg \varepsilon$ (in case of

SOU process) are not useful for obtaining accurate estimates when ε is fixed and unknown. Assuming N is large enough so that its contribution to the bias is negligible, the graph of the estimators $\hat{\gamma}_\varepsilon = \hat{\gamma}_\varepsilon(\Delta)$ and $\hat{\sigma}^2_\varepsilon = \hat{\sigma}^2_\varepsilon(\Delta)$ do not indicate the optimal regime for subsampling. It is evident from these graphs that the bias is inversely proportional to the time-step Δ, but no cut-offs or critical values can be inferred.

However, using the estimators evaluated at multiple values of $h = \Delta_1, \ldots, \Delta_M$ (by subsampling the dataset), and then, estimating the slope of the best-fit straight line to $\Gamma(h) = h \times \hat{\gamma}_\varepsilon(h)$ gives a more accurate estimate of the parameter γ (hence, σ^2). The value h at which there is a curvature change in the curve $\Gamma(h)$ for h close to zero, provides an estimate for the unknown small parameter ε, and identifies the optimal subsampling regime.

This methodology is numerically verified for the additive triad model, and the truncated Burgers-Hopf model. It is conjectured that this methodolgy is generic and can be used for estimation under indirect observability in more complex settings, with no knowledge of the small parameter ε, and will also help identify the degree of approximation of the observed dataset, as compared to the underlying unobservable process.

Chapter 6

Generic Indirect Observability Framework

6.1 Introduction

A general framework of *indirect observability* is introduced in this chapter. In the preceding chapters we have successfully demonstrated the validity of various optimal subsampling schemes to achieve asymptotic consistency of the estimators in particular for the OU parameters γ, σ^2. The discrete observations for estimation were subsampled from an approximating process; the Smoothed OU process as one example, and the multiscale dynamics given by the additive triad model, truncated Burgers-Hopf model, as a second example. The asymptotic study of the parameters is carried in two major steps. First, the asymptotic properties of the *empirical covariance estimators* $\hat{r}_k^\epsilon$ is understood through the computations of their L_2-norms. Secondly, the relation between the OU parameters and the empirical covariance estimators is exploited to transfer the asymptotic results obtained for the estimators $\hat{r}_k^\epsilon$ to the OU estimators. This straightforward two-step analysis is extended to a more general class of processes.

Formally, consider a continuous-time centered stationary random process $X = \{X_t\}$ in L_2, with covariance function $K(u) = E[X_t X_{t+u}]$ defined by a parametric model $K(u) = K(u, \theta)$, where $\theta \in \Theta$ is an unknown vector of parameters, and Θ is a bounded open subset of $\mathbb{R}^p$. We focus here on situations where the stationary process $X = \{X_t\}$ is not directly observable, and where the only available observations are generated by centered stationary processes $Y^\epsilon = \{Y_t^\epsilon\}$ indexed by $\varepsilon > 0$, such that Y^ϵ tends to X in some adequate sense as $\varepsilon \to 0$. The concrete target is to efficiently use these approximate data to generate consistent estimators of the unknown "underlying" parameter vector θ.

In chapters 3 and 4, we presented several such cases where it is essential to first implement an adequate subsampling $\{Y^\epsilon_{n\Delta}, 1 \leq n\}$ of the approximate process Y_t^ϵ, using a small time interval $\Delta = \Delta(\epsilon) \to 0$, and a number of observations $N(\varepsilon) \to \infty$, as $\epsilon \to 0$. For instance, in chapter 3, this high density subsampling scheme is applied to purely Gaussian processes Y^ϵ and X. Specifically in these cases, the associated adaptive subsampling schemes enabling asymptotically consistent estimation of θ on the basis of indirect observations, are explicitly characterized.

Now, we analyze here a much more general situation covering a wide range of applications, namely, the approximating stationary processes Y_t^ϵ have *arbitrary probability distributions* with uniformly bounded fourth-order moments, and the limiting process X_t is Gaussian. We characterize the efficient pairings of subsampling rates $\Delta = \Delta(\epsilon)$ and number of observations $N = N(\epsilon)$ which generate consistent estimators of the unknown parameters θ. As seen below, a key technical point is to obtain consistent estimators of covariances, which in turn generate consistent estimators $\hat{\theta}_\varepsilon$ of θ. For instance, representation of the OU estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}^2_\varepsilon$ in terms of the standard empirical covariance estimators $\hat{r}^\varepsilon_0 = \hat{r}^\varepsilon_0(N(\varepsilon), \Delta(\varepsilon))$, and $\hat{r}^\varepsilon_1 = \hat{r}^\varepsilon_1(N(\varepsilon), \Delta(\varepsilon))$, namely,

$$\hat{\gamma}_\varepsilon = (1/\Delta) \ln(\hat{r}^\varepsilon_0 / \hat{r}^\varepsilon_1), \qquad \hat{\sigma}^2_\varepsilon = 2\hat{\gamma}_\varepsilon \hat{r}^\varepsilon_0, \tag{6.1}$$

has been crucial in deducing their asymptotic consistency under indirect observability.

An important notion used in the general framework is that of defining the empirical covariance estimators with *non-vanishing* lags, defined below. Its equivalence in the specific case of the OU process is that the estimators (6.1) are defined with Δ replaced with a continuous time lag $u > 0$, which remains bounded away from zero under the asymptotic conditions.

The specific example of the *smoothed Ornstein-Uhlenbeck* process presented in chapter 3 is revisited with the results from the general setting. Given discrete observations subsampled from the *smoothed Ornstein-Uhlenbeck* process, the estimators with non-vanishing lags, introduced in the general framework, provide higher order of accuracy for the estimation of OU parameters.

6.2 Indirect Observability Context

Let us describe the precise mathematical framework studied in this chapter, which we will call the *Indirect Observability Context* for the centered stationary continuous time processes Y^ε and X.

Hypotheses on the unobservable process X

We call $\mathbb{G}$ the set of all continuous-time centered stationary Gaussian processes $X = \{X_t\}$ with covariance function $K(u, \theta)$ parametrized by the vector $\theta \in \Theta$ where Θ is an open bounded subset of $\mathbb{R}^p$. The covariance function $K(u, \theta)$ is assumed to be piecewise C^1 in the time lag u and C^1 in θ, and to decay exponentially fast for large time lags, i.e.,

$$|K(u, \theta)| < c \exp(-|u|/a), \quad \forall u \in \mathbb{R}, \quad \forall \theta \in \Theta,$$

where the *decay coefficients* c and a are arbitrary but fixed positive constants. The underlying process $X = \{X_t\}$ is *not directly observable*, and will always be assumed to belong to the preceding class $\mathbb{G}$ of stationary Gaussian processes.

Hypotheses on the observable approximating processes Y^ε

For each fixed $\epsilon > 0$, the only observations available are the approximate data Y_t^ϵ. The observable process $Y^\varepsilon = \{Y_t^\varepsilon\}$ is assumed to be centered stationary, but may have arbitrary probability distributions. We assume that the random variables Y_t^ε *converge in* L_4 to X_t as $\varepsilon \to 0$. More precisely we assume that there is a fixed continuous function $\rho(\varepsilon) > 0$ tending to 0 as $\varepsilon \to 0$ and a fixed positive constant C such that

$$\|Y_t^\varepsilon\|_{L_4} < C, \quad \text{and} \quad \|Y_t^\varepsilon - X_t\|_{L_4} < \rho(\varepsilon), \qquad \forall t \in \mathbb{R},\ \forall \varepsilon > 0,\ \forall \theta \in \Theta.$$

Our goal is to use the approximate data Y^ϵ to compute consistent estimators of the unknown parameter θ. The uniform L_4 - speed of convergence $\rho(\varepsilon)$ of Y_t^ε to X_t will naturally become below a critical characteristic of the subsampling rates enabling asymptotic consistency of parameter estimators.

As proved below in proposition 6.5.1, the L_2-norm of the difference between the empirical covariance estimators associated to two distinct processes is bounded above by the L_4-norm of the difference between the processes, which justifies the above hypotheses.

Adaptive Subsampling schemes

Under these *Indirect Observability* hypotheses, we systematically apply to the observed process Y^ε one or several *adaptive subsampling* schemes defined by $\Delta = \Delta(\varepsilon)$, and $N = N(\varepsilon)$ such that

$$\Delta(\varepsilon) \to 0, \quad \text{and } N(\varepsilon)\Delta(\varepsilon) \to \infty \quad \text{as } \varepsilon \to 0. \tag{6.2}$$

Namely each such sampling scheme will focus on the observed subsamples $Y^{\varepsilon}_{n\Delta}$ with $n = 1 \dots N(\varepsilon)+k(\varepsilon)$, where the discrete time lags $k(\varepsilon)$ are integer valued functions of ε such that $k(\varepsilon)\Delta(\varepsilon)$ tend to some finite limit as $\varepsilon \to 0$. The main goal of such subsampling schemes is to enable consistent estimation of the covariances $K(u, \theta)$.

6.3 Parameter Estimators and Covariance Estimators

Since centered Gaussian processes with the same covariance functions are probabilistically equivalent, it is quite natural to assume that the covariance function $u \to K(u, \theta)$ determines θ uniquely. A slightly stronger hypothesis is to assume the existence of a finite fixed set of time lags $\{u_1, ..., u_p\}$ such that the vector $r = (r_1, \dots, r_p)$ of covariances $r_j = K(u_j, \theta)$ determines a unique $\theta \in \Theta$ denoted by $\theta = F(r)$. In this case, a few mild hypotheses imply that consistent estimation for parameters or for covariances are equivalent tasks, as indicated by the following proposition.

Proposition 6.3.1 *Let $X = \{X_t\}$ be a centered stationary random process with covariance function $K(u) = K(u, \theta)$ parametrized by $\theta \in \Theta \subset \mathbb{R}^p$. Assume that the function $K(u, \theta)$ is continuously differentiable with respect to θ and piecewise C^1 with respect to u. Select and fix a set of p time lag values $\{u_1, ..., u_p\}$, and consider the system of p equations with p unknowns*

$$K(u_j, \theta) = r_j, \quad j = 1...p. \tag{6.3}$$

Assume that the Jacobian determinant $J(\theta)$ is non-zero for all $\theta \in \Theta$, where

$$J(\theta) = det \left[\left(\frac{\partial K(u_j, \theta)}{\partial \theta_k} \right)_{1 \le j,k \le p} \right].$$

The set $\mathbb{D} \subset \mathbb{R}^p$ of all vectors $r = (r_1, \dots, r_p)$ of the form (6.3) for some $\theta \in \Theta$ is then open. Assume that for all $r \in \mathbb{D}$ the solution of system (6.3) is unique in Θ, and denote this

solution by $\theta = F(r)$. Consider an arbitrary family of observed processes $Y^\varepsilon = \{Y_t^\varepsilon\}$ indexed by $\varepsilon > 0$. Then as $\varepsilon \to 0$, there exists a family of asymptotically consistent estimators $\hat{\theta}(\varepsilon)$ of θ based on Y^ε if and only if there exists a family of asymptotically consistent estimators $\hat{r}(\varepsilon)$ of r based on Y^ε. One may for instance link these two families of estimators by $\hat{\theta}(N) = F(\hat{r}(N))$ and $K(u_j, \hat{\theta}(\varepsilon)) = \hat{r}_j(\varepsilon)$. Moreover, the estimators $\hat{r}(\varepsilon)$ are asymptotically Gaussian if and only if the estimators $\hat{\theta}(\varepsilon)$ are also asymptotically Gaussian.

Proof. By the implicit function theorem [40], the function $F(r)$ is continuously differentiable in r, for $r \in \mathbb{D}$. Then the proof is quite classical, see for instance [10]. ■

In our context this proposition will be systematically applied to empirical covariance estimators based on processes Y^ε approximating X and subsampled by adaptive subsampling schemes (6.2).

6.4 Subsampled Empirical Covariances

We now present empirical covariance estimators based on observations subsampled from the approximating process Y^ε. The subsampling time interval $\Delta = \Delta(\varepsilon)$, and number of observations $N = N(\varepsilon)$ will be functions of ε verifying (6.2). We select integer valued covariance lags $k = k(\varepsilon)$ depending on ε, and we define the associated *approximate subsampled empirical covariances* $K_Y^\varepsilon = \hat{r}_k(N, \Delta)$ computed from $N + k$ subsampled observations $V_n = Y^\varepsilon_{n\Delta(\varepsilon)}$ by the formula

$$K_Y^\varepsilon = \hat{r}_k(N, \Delta) = \frac{1}{N} \sum_{n=1}^{N} V_n V_{n+k} = \frac{1}{N} \sum_{n=1}^{N} Y^\varepsilon_{n\Delta} Y^\varepsilon_{(n+k)\Delta}. \tag{6.4}$$

Similarly, we define the *true* subsampled empirical covariances K_X^ε based on the process X by

$$K_X^\varepsilon = \frac{1}{N} \sum_{n=1}^{N} U_n U_{n+k} = \frac{1}{N} \sum_{n=1}^{N} X_{n\Delta} X_{(n+k)\Delta}, \tag{6.5}$$

where $U_n = X_{n\Delta}$ is the subsampled *unobserved* process X_t. Note that for the true subsampled empirical covariances K_X^ε the indexing parameter ε is only a *mute* variable. The direct observations from the unobserved process X_t are considered on the same time grid defined by ε, N, Δ for comparison with the *approximate subsampled empirical covariances* K_Y^ε. For each fixed time lag $u > 0$, we want to estimate the covariance $K(u, \theta)$ by the approximate subsampled empirical covariance K_Y^ε . It is then natural to select integer valued covariance lags $k(\varepsilon)$ such that $k(\varepsilon)\Delta(\varepsilon) \to u$ as $\varepsilon \to 0$. It turns out (see below in section 6.7) that the optimal adaptive schemes are obtained when $k(\varepsilon)\Delta(\varepsilon) = u$, for all $\varepsilon > 0$. To be in this optimal situation, we first select the numbers of observations $N(\varepsilon)$ and integer valued covariance lags $k(\varepsilon)$ such that as $\varepsilon \to 0$,

$$N(\varepsilon) \to \infty, \qquad k(\varepsilon) \to \infty, \qquad k(\varepsilon)/N(\varepsilon) \to 0, \tag{6.6}$$

and we then define the subsampling time interval by

$$\Delta(\varepsilon) = u/k(\varepsilon) \quad \text{for each fixed time lag} \quad u > 0. \tag{6.7}$$

For the particular time lag $u = 0$, where we want to estimate the variance $K(0, \theta)$, we of course impose $k(\varepsilon) = 0$ for all $\varepsilon > 0$, but no other restriction on $\Delta(\varepsilon)$; the only requirements on $N(\varepsilon), \Delta(\varepsilon)$ are that they must verify (6.2). To estimate p parameters we need typically to estimate p distinct correlations $K(u_j, \theta)$, including the variance of X_t corresponding to $u_1 = 0$. Therefore, an optimized approach will require the selection of $p - 1$ distinct subsampling time intervals $\Delta_2, \ldots, \Delta_p$. When some of the ratios u_i/u_j are integers, with $u_i > 0$, $u_j > 0$, it is clearly possible to select the same time intervals $\Delta_i = \Delta_j$ provided one picks adequate distinct covariance lags $k(\varepsilon)$ to estimate $K(u_i)$ and $K(u_j)$. We will say that the K_Y^ε are empirical covariance estimators with *non-vanishing lags*, since the integers $k(\varepsilon)$ tend to ∞.

The *vanishing lags* asymptotics (see section 6.9) correspond to time lags of the form $u = u(\varepsilon) = k\Delta(\varepsilon)$ where the discrete lag k is any fixed integer such that $u(\varepsilon) \to 0$ as $\varepsilon \to 0$. The vanishing lags asymptotics for the general framework will be studied elsewhere.

In the following sections it is proved that for adequate choices of the adaptive subsampling schemes $\Delta(\varepsilon), N(\varepsilon)$ the approximate subsampled empirical covariances K_Y^ε with non-vanishing lags converge in L_2 to $K(u,\theta)$ as $\varepsilon \to 0$. We will often omit the argument θ in the covariances $K(u) = K(u,\theta)$. But we point out that all bounds and constants derived below are uniform when θ remains in any fixed compact subset of Θ.

6.5 Impact of Approximate Data on Empirical Covariances

First we derive an upper bound for the L_2 distance between the *"approximate"* empirical covariances K_Y^ε based on Y^ε defined in (6.4) and the *"true"* empirical covariances K_X^ε defined in (6.5). Denote the L_p-norm of a random variable Z by $\|Z\|_{L_p}$.

Proposition 6.5.1 *Consider two discrete centered stationary processes $U = \{U_n\}$, and $V = \{V_n\}$ in L_4. Assume that for some positive constants M and $d > 0$, the following inequalities hold in L_4 for all integers n,*

$$\|V_n - U_n\|_{L_4} \le M, \qquad \|U_n\|_{L_4} \le d \quad , and \quad \|V_n\|_{L_4} \le d. \tag{6.8}$$

Consider the usual empirical covariance estimators of processes U and V at discrete time lags k defined by,

$$\hat{r}_U(k,N) = (1/N)\sum_{n=1}^{N} U_n U_{n+k}, \qquad \hat{r}_V(k,N) = (1/N)\sum_{n=1}^{N} V_n V_{n+k}.$$

We then have the uniform majoration in L_2-norm

$$\|\hat{r}_V(k,N) - \hat{r}_U(k,N)\|_{L_2} \leq 2dM, \quad \textit{for all } N \textit{ and } k.$$

Proof. We first prove a key lemma.

Lemma 6.5.2 *Consider 4 random variables A, A', B, B' in L_4. Then, we have*

$$\|AB - A'B'\|_{L_2} \leq \|A - A'\|_{L_4}\|B\|_{L_4} + \|A'\|_{L_4}\|B - B'\|_{L_4},$$

Proof. For any pair of random variables Q, R in L_4, we have by Cauchy-Schwarz inequality,

$$\|QR\|_{L_2} = \sqrt{E(Q^2R^2)} \leq \sqrt{[E(Q^4)]^{1/2}[E(R^4)]^{1/2}} = \|Q\|_{L_4}\|R\|_{L_4}. \tag{6.9}$$

The elementary bound,

$$\|AB - A'B'\|_{L_2} = \|(A - A')Y + A'(B - B')\|_{L_2} \leq \|(A - A')B\|_{L_2} + \|A'(B - B')\|_{L_2},$$

combined with inequality (6.9) proves the lemma. ■

We now prove proposition 6.5.1. The hypotheses of this proposition and the preceding lemma imply

$$\|V_nV_{n+k} - U_nU_{n+k}\|_{L_2} \leq \|V_n - U_n\|_{L_4}\|V_{n+k}\|_{L_4} + \|U_n\|_{L_4}\|V_{n+k} - U_{n+k}\|_{L_4} \leq 2dM.$$

From this inequality we derive for all N and k

$$\|\hat{r}_{k,V}(N) - \hat{r}_{k,U}(N)\|_{L_2} \leq (1/N)\sum_{n=1}^{N}\|V_nV_{n+k} - U_nU_{n+k}\|_{L_2} \leq 2dM.$$

■

We now derive a crucial consequence for the indirect observability context.

Theorem 2 *Consider processes Y^ε and X verifying the Indirect Observability hypotheses of section 6.2. Call C the uniform bound of the $\|Y_t^\varepsilon\|_{L_4}$ and $\rho(\varepsilon)$ the uniform L_4-speed of convergence of the Y_t^ε to X_t. Let σ^2 be a fixed upper bound for the variance $K(0)$ of X_t.*

For arbitrary $\varepsilon > 0$, subsampling rate $\Delta(\varepsilon)$, number of observations $N(\varepsilon)$, and discrete covariance lags $k(\varepsilon)$, consider the approximate covariance estimators K_Y^ε based on Y^ε and the true covariance estimators K_X^ε based on X, respectively given by (6.4) *and* (6.5)*.*

Then, for all ε we have the following uniform L_2 -bound for the difference between true and approximate covariance estimators

$$\|K_Y^\varepsilon - K_X^\varepsilon\|_{L_2} \leq 2d\ \rho(\varepsilon), \tag{6.10}$$

where $d = \max\{C, (3\sigma/2)\}$.

Proof. Since X is Gaussian, the subsampled process $U_n = X_{n\Delta}$ satisfies,

$$\|U_n\|_{L_4} = \|X_{n\Delta}\|_{L_4} = 3^{1/4}\sigma < \frac{3}{2}\sigma.$$

The Indirect Observability hypotheses on Y_t^ϵ imply that the process $V_n = Y_{n\Delta}^\varepsilon$ satisfies

$$\|V_n\|_{L_4} \leq C, \quad \text{and} \quad \|V_n - U_n\|_{L_4} \leq \rho(\varepsilon).$$

Then, proposition 6.5.1 implies the announced inequality (6.10). ■

The preceding theorem shows that the asymptotic behavior in L_2 of the approximate empirical covariances K_Y^ε as $\varepsilon \to 0$ will be determined by the comparison of the function $\rho(\varepsilon)$ with the asymptotic behavior of the true empirical covariances K_X^ε. Hence, we focus on the precise asymptotics of subsampled empirical covariances K_X^ε for Gaussian processes X.

6.6 Gaussian Processes : Accuracy Bounds for Covariance Estimators

We first consider discrete Gaussian processes with no subsampling, and compute accuracy bounds in L_2 for empirical covariance estimators. Recall a known result ([10] for instance).

Proposition 6.6.1 *Let $U = \{U_n\}$ be a centered stationary Gaussian process with covariances $r_k = E[U_n U_{n+k}]$. Denote the empirical covariance estimators based on $N + k$ observations by*

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

Call $\Gamma = \Gamma(N)$ the covariance matrix of these estimators, given by

$$\Gamma_{k,q} = E\left[(\hat{r}_k - r_k)(\hat{r}_q - r_q)\right], \quad \forall k \geq 0,\ q \geq 0.$$

Then, for all pairs k, q of non-negative integers one has

$$\Gamma_{k,q} = (1/N) \left[\sum_{j=-(N-1)}^{N-1} f(j) \right] - (1/N^2) \left[\sum_{j=1}^{N-1} j(f(j) + f(-j)) \right],$$

where the $f(j) = f_{k,q}(j)$ are defined by the covariances of U as follows,

$$f(j) = \left(r_j r_{j+q-k} + r_{j+q} r_{j-k}\right). \tag{6.11}$$

Proof. Fix arbitrary non-negative integers k, q. The 2nd moments of empirical covariance estimators are given by,

$$N^2 E[\hat{r}_k \hat{r}_q] = \sum_{n=1}^{N} \sum_{p=1}^{N} E[U_n U_{n+k} U_p U_{p+q}].$$

U is Gaussian, therefore, the fourth-moments can be expressed in terms of second-moments, and we obtain,

$$E[U_n U_{n+k} U_p U_{p+q}] = r_k r_q + f(p-n),$$

where we define the numbers $f(j)$ as above in (6.11). Hence, the covariances of the empirical estimators are given by

$$\Gamma_{k,q}(N) = E[\hat{r}_k \hat{r}_q] - r_k r_q = (1/N^2) \sum_{n=1}^{N} \sum_{p=1}^{N} f(p-n).$$

For any function f defined on the the set of integers, we have the identity

$$\frac{1}{N^2} \sum_{n=1}^{N} \sum_{p=1}^{N} f(p-n) = \frac{1}{N} \left[\sum_{j=-(N-1)}^{N-1} f(j) \right] - \frac{1}{N^2} \left[\sum_{j=1}^{N-1} j(f(j) + f(-j)) \right]. \tag{6.12}$$

This identity achieves the proof. ■

We now evaluate the L_2-accuracies $\|\hat{r}_k - r_k\|_{L_2}$ of empirical covariance operators.

Proposition 6.6.2 *Consider a centered stationary Gaussian process $U = \{U_n\}$, with covariances r_j. Define τ and ζ by*

$$\tau = \sum_{j \geq 1} j r_j^2, \qquad \zeta = \sum_{j \in Z} r_j^2, \tag{6.13}$$

and assume that τ is finite, which clearly implies that ζ is finite. Then the L_2 accuracy of the empirical covariances is uniformly bounded by

$$\|\hat{r}_k(N) - r_k\|_{L_2} \ \leq \ \frac{\sqrt{2\zeta}}{\sqrt{N}} + \frac{\left(2\tau^{1/2} + (4\tau\zeta k)^{1/4}\right)}{N}, \quad \forall N, k \geq 0. \tag{6.14}$$

In particular, for each $k \geq 0$, as $N \to \infty$, the real-valued empirical covariance estimator $\hat{r}_k = \hat{r}_k(N)$, converges in L_2 towards the true covariance r_k with an L_2-speed of convergence

given by,

$$\lim_{N\to\infty} N\|\hat{r}_k(N) - r_k\|_{L_2}^2 = \sum_{j\in Z} \left(r_j^2 + r_{j+k}r_{j-k}\right). \tag{6.15}$$

Proof. Since, $r_k = r_{-k}$, we may assume $k \geq 0$. Proposition 6.6.1 shows that the variance of $\hat{r}_k$ is given by $\Gamma_{k,k} = S/N + R(N)$, where

$$S = S(k) = \sum_{j\in\mathbb{Z}} f(j), \tag{6.16}$$

and

$$R = -(2/N)\left(\sum_{j\geq N} f(j)\right) - (2/N^2)\left(\sum_{j=1}^{N-1} jf(j)\right),$$

with $f(j) = f(-j) = r_j^2 + (r_{j-k}r_{j+k})$. By Cauchy-Schwarz inequality we then have $|S| \leq 2\zeta$. Since $N|f(j)| \leq j|f(j)|$ when $j \geq N$, we obtain

$$|R| \leq \left(\frac{2}{N^2}\right)\sum_{j\geq 1} j|r_j^2 + r_{j-k}r_{j+k}| \leq \left(\frac{2}{N^2}\right)\left(\tau + \sum_{j\geq 1} j\,|r_{j-k}r_{j+k}|\right).$$

By Cauchy-Schwarz inequality, we have

$$\left(\sum_{j\geq 1} j|r_{j+k}||r_{j-k}|\right) \leq \left(\sum_{j\geq 1} jr_{j+k}^2\right)^{1/2}\left(\sum_{j\geq 1} jr_{j-k}^2\right)^{1/2}.$$

To bound the first factor in the last inequality, note that

$$\sum_{j\geq 1} jr_{j+k}^2 \leq \sum_{j\geq 1}(j+k)r_{j+k}^2 \leq \tau,$$

where τ is given by (6.13). To bound the second factor, we write

$$\sum_{j\geq 1} jr_{j-k}^2 = \sum_{j\geq k}(j-k)r_{j-k}^2 + k\sum_{j\geq k} r_{j-k}^2 + \sum_{j=1}^{k-1} jr_{j-k}^2 \leq \tau + \zeta k.$$

We, thus, obtain the bound

$$|R| \leq \frac{2\left(\tau + \sqrt{\tau^2 + \tau\zeta k}\right)}{N^2} \leq \frac{\left(4\tau + 2\sqrt{\tau\zeta k}\right)}{N^2},$$

and hence,

$$\sqrt{|R|} \leq \frac{\left(2\tau^{1/2} + (4\tau\zeta k)^{1/4}\right)}{N}.$$

We, then, have

$$\|\hat{r}_k - r_k\|_{L_2} = \Gamma_{k,k}^{1/2} = (S/N + |R|)^{1/2} \leq (S/N)^{1/2} + |R|^{1/2},$$

and therefore,

$$\|\hat{r}_k - r_k\|_{L_2} \leq \frac{\sqrt{2\zeta}}{\sqrt{N}} + \frac{\left(2\tau^{1/2} + (4\tau\zeta k)^{1/4}\right)}{N}.$$

This proves the announced accuracy bound (6.14), and as an obvious consequence, the asymptotic result (6.15). ■

We now extend the preceding accuracy bounds to the subsampled empirical covariances of Gaussian processes $X \in \mathbb{G}$.

Theorem 3 *Consider a centered stationary Gaussian process X belonging to $\mathbb{G}$ (see (6.2)). To estimate the covariance $K(u)$ by subsampled empirical covariances K_X^{ε}, we distinguish two cases (a) and (b):*

Case (a) : For any fixed strictly positive time lag u, we select the numbers of observations $N(\varepsilon) \to \infty$ and integer valued covariance lags $k(\varepsilon) \to \infty$ such that $k(\varepsilon)/N(\varepsilon) \to 0$, and then we define the associated subsampling time intervals by $\Delta(\varepsilon) = u/k(\varepsilon)$. One has then $\Delta(\varepsilon) \to 0$ and $N(\varepsilon)\Delta(\varepsilon) \to \infty$.

Case (b) : For the particular time lag $u = 0$, we select numbers of observations $N(\varepsilon) \to \infty$ and subsampling time intervals $\Delta(\varepsilon) \to 0$ such that $N(\varepsilon)\Delta(\varepsilon) \to \infty$, but we impose $k(\varepsilon) = 0$ for all $\varepsilon > 0$.

In both of these cases we thus impose $k(\varepsilon)\Delta(\varepsilon)$ identically equal to u, and we consider the associated subsampled empirical covariances K_X^ε, given by (6.5), based on $(N(\varepsilon)+k(\varepsilon))$ observations $U_n = X_{n\Delta(\varepsilon)}$.

Then, the subsampled empirical covariances K_X^ε converge in L_2 to the true covariance $K(u)$ as $\varepsilon \to 0$. For u in any fixed interval $[0, J]$, the L_2-speed of convergence is given by the uniform bound

$$\|K_X^\varepsilon - K(u)\|_{L_2} \leq \ cte/\sqrt{N(\varepsilon)\Delta(\varepsilon)}, \quad \forall \varepsilon < 1/cte. \tag{6.17}$$

Moreover, the precise L_2-speed of convergence is given by,

$$\lim_{\varepsilon\to 0} N(\varepsilon)\Delta(\varepsilon)\|K_X^\varepsilon - K(u)\|_{L_2}^2 = \int_{-\infty}^{+\infty} ds \left(K(s)^2 + K(s+u)K(s-u)\right). \tag{6.18}$$

In this accuracy estimate, "cte" denotes a positive constant depending only on J and on the exponential decay coefficients c, a associated to the covariance function $K(u)$ of X. Note also that for $u > 0$ this speed of convergence is inversely proportional to u, since Δ is proportional to u.

Proof. For brevity, we omit the ε argument in N, k, Δ. Since, X is in $\mathbb{G}$, call c and a the coefficients of exponential decay of its covariances. The discrete Gaussian process $U_n = X_{n\Delta(\varepsilon)}$ has, then, covariances r_j bounded, namely, $r_j = K(j\Delta(\varepsilon)) \leq c\exp(-j\Delta/a), \quad \forall j \geq 0$. The coefficients τ, ζ introduced by (6.13) verify the following bounds valid for $\Delta(\varepsilon) < a/4$,

$$\begin{aligned} \tau &\leq c\sum_{j\leq 1} j\exp(-2j\Delta/a) \leq 4ca^2/\Delta^2, \\ \zeta &\leq 2c^2\sum_{j\geq 0}\exp(-2j\Delta/a) \leq 2c^2 a/\Delta. \end{aligned} \tag{6.19}$$

We first study the case of a fixed positive time lag $u > 0$, where $\Delta = u/k$. The preceding inequalities imply, for $4u/a < k(\varepsilon)$,

$$\begin{aligned} \tau &\leq 4ca^2k^2/u^2, \\ \zeta &\leq 2c^2ak/u. \end{aligned}$$

Then, the covariance estimator K_X^ε generated by subsampling X coincides with the empirical covariance estimator $\hat{r}_{k(\varepsilon)}$ based on $N+k$ observations of the discrete Gaussian process U. The corresponding exact covariance of U is $r_k = K(k\Delta) = K(u)$. Applying inequality (6.14) for the process U and the integer valued covariance lag $k = k(\varepsilon)$, we have the inequality

$$\|K_X^\varepsilon - K(u)\|_{L_2} \leq \frac{\sqrt{2\zeta}}{\sqrt{N}} + \frac{\left(2\tau^{1/2} + (4\tau\zeta k)^{1/4}\right)}{N}. \tag{6.20}$$

Injecting the bounds just computed for τ and ζ, we obtain

$$\|K_X^\varepsilon - K(u)\|_{L_2} \leq 2c\sqrt{a}\sqrt{k/uN} + \left(4c^{1/2}a + 2^{5/4}(ca)^{3/4}u^{1/4}\right)(k/uN). \tag{6.21}$$

This obviously implies the desired bound

$$\|K_X^\varepsilon - K(u)\|_{L_2} \leq \text{cte}\sqrt{k/uN} = \text{cte}/\sqrt{N\Delta},$$

where the constant "cte" depends only on c, a, and ε is small enough to force $k/N < u$ and $k > 4u/a$. In particular, this restriction on ε can be handled uniformly for $0 < \alpha \leq u \leq \beta$ with fixed α, β. We now study the particular case $u = 0$, where one seeks to estimate the variance $K(0)$. We impose $k = k(\varepsilon) = 0$, and the subsampling scheme N, Δ verifies (6.2) as stated in the theorem. Then, (6.20) for $u = 0$ is specifically given as,

$$\|K_X^\varepsilon - K(0)\|_{L_2} \leq \frac{\sqrt{2\zeta}}{\sqrt{N}} + \frac{\left(2\tau^{1/2}\right)}{N}.$$

Using (6.19) we obtain for $\Delta < a/4$,

$$\tau \leq 4ca^2/\Delta^2, \qquad \zeta \leq 2c^2 a/\Delta,$$

and we conclude,

$$\|K_X^\varepsilon - K(0)\|_{L_2} \leq \frac{2c\sqrt{a}}{\sqrt{N\Delta}} + \frac{4a\sqrt{c}}{N\Delta},$$

which implies, as soon as ε is small enough to force $N\Delta > 1$ and $\Delta < a/4$,

$$\|K_X^\varepsilon - K(0)\|_{L_2} \leq \text{cte}/\sqrt{N\Delta}.$$

To compute the precise speed, given by (6.18), of convergence for $\|K_X^\varepsilon - K(0)\|_{L_2}$ to zero, one simply needs to implement a variant of the result (6.15), proved above, for discrete processes. ■

Conclusions for directly observable Gaussian processes

It can be shown that, under the Gaussian hypotheses of the preceding result, the speed of convergence $1/\sqrt{N\Delta}$ obtained above is optimal (up to a proportionality constant) for subsampled covariance estimators based on adaptive subsampling schemes $N(\varepsilon)$, $\Delta(\varepsilon)$ verifying (6.2). Hence, when a Gaussian process X_t with exponentially decaying covariances is *directly observable*, basic adaptive subsampling schemes $\Delta(\varepsilon), N(\varepsilon)$ verifying (6.2) actually generate subsampled empirical covariances K_X^ε which are L_2-consistent estimators of true covariances $K(u)$. Moreover, as $\varepsilon \to 0$, provided the integer time lags $k(\varepsilon)$ and subsampling time intervals $\Delta(\varepsilon)$ are selected such that $k\Delta = u$, the estimators achieve the optimal L_2-speed of convergence $1/\sqrt{N\Delta}$ up to a proportionality constant.

The $(N + k)$ subsampled observations used to estimate $K(u)$ cover a global observation time span $Span(\varepsilon) = (N + k)\Delta$ for the process X_t, and clearly, $\lim_{\varepsilon \to 0}(Span/N\Delta) = 1$. Hence, the L_2-accuracy of the subsampled covariance estimators is proportional to

$1/\sqrt{Span}$, as could be expected.

We also remark that, for $u > 0$, one could have proposed a slightly different strategy to generate subsampled covariance estimators, specifically, first selecting an adaptive subsampling scheme $N(\varepsilon)$, $\Delta(\varepsilon)$ verifying (6.2), and then, defining a covariance estimator K_X^ε based on integers $k(\varepsilon) = [u/\Delta(\varepsilon)]$, where $[x]$ is the integer closest to x. For this strategy, a proof similar to the preceding one yields a slightly less good L_2-speed of convergence estimate. It involves a local bound $\chi(u) = sup_{w\in[0,u]}|K'(w)|$ on the first derivative of the covariance function, namely,

$$\|K_X^\varepsilon - K(u)\|_{L_2} \;\leq\; \chi(u)\Delta \;+\; \frac{\text{cte}}{\sqrt{N\Delta}}\,, \tag{6.22}$$

where the constant "cte" depends only on c, a, $\chi(u)$, u, and remains bounded when u is bounded. This alternative strategy is less favorable than the strategy outlined in the preceding theorem. Indeed, for a given $N(\varepsilon)$, the best reachable L_2-speed of convergence is obtained by minimizing the right hand side of (6.22); this obviously occurs when Δ^3 is proportional to $1/N$, and the optimal accuracy reachable by this strategy is, then, proportional to $\Delta = cte/\sqrt{N\Delta} = cte/N^{1/3}$. We now come back to the generic Indirect Observability context.

6.7 Accuracy of Subsampled Approximate Empirical Covariance Estimators

Theorem 4 *Consider centered stationary observable processes Y_t^ε with arbitrary probability distributions. We assume that as $\varepsilon \to 0$, the Y_t^ε converge in L_4 to an unobservable stationary Gaussian process X_t, and verify all the Indirect Observability hypotheses of section 6.2. Call C a constant bound for the L^4-norms of all the Y_t^ε. Let $\rho(\varepsilon) > 0$ tending to 0 as $\varepsilon \to 0$ be a bound (uniform in t) for all the L_4-norms $\|Y_t^\varepsilon - X_t\|_{L_4}$. Let σ^2 be an upper bound for the*

variance of X_t for all θ in some fixed compact subset Γ of Θ.

To estimate the covariance $K(u)$ of X by subsampled empirical covariances K_Y^ε based on the observable approximate data Y^ε we distinguish two cases :

Case (a) (positive covariance lags) : Select the numbers of observations $N(\varepsilon) \to \infty$ and the integer valued covariance lags $k(\varepsilon) \to \infty$ such that $k(\varepsilon)/N(\varepsilon) \to 0$. Then for each fixed strictly positive covariance time lag u, define associated subsampling time intervals by $\Delta(\varepsilon) = u/k(\varepsilon)$. One has then $\Delta(\varepsilon) \to 0$ and $N(\varepsilon)\Delta(\varepsilon) \to \infty$.

Case (b) (zero covariance lag) : For the particular covariance time lag $u = 0$, we select numbers of observations $N(\varepsilon) \to \infty$, and subsampling time intervals $\Delta(\varepsilon) \to 0$ such that $N(\varepsilon)\Delta(\varepsilon) \to \infty$, but we naturally impose $k(\varepsilon) = 0$ for all $\varepsilon > 0$.

In both of these cases, we thus impose $k(\varepsilon)\Delta(\varepsilon)$ identically equal to u and we consider the corresponding approximate subsampled empirical covariances K_Y^ε, given by (6.4), based on $(N(\varepsilon) + k(\varepsilon))$ observations $V_n = Y_{n\Delta(\varepsilon)}$.

Then, as $\varepsilon \to 0$, the approximate subsampled empirical covariances K_Y^ε converge in L_2 to the true covariance $K(u)$ of the process X. For $0 \leq u < J$ with J fixed, the L_2-speed of convergence is given by the uniform bound

$$\|K_Y^\varepsilon - K(u)\|_{L_2} \leq 2d\rho(\varepsilon) + cte\,/\sqrt{N(\varepsilon)\Delta(\varepsilon)}, \quad \forall \varepsilon < 1/cte, \tag{6.23}$$

where the constant $d = \max\{C, (3\sigma/2)\}$.

Here and below, "cte" denotes a positive constant depending only on the time interval J and on the exponential decay coefficients c, a associated to the covariance function K of X.

The optimal L_2-speed of convergence for $\|K_Y^\varepsilon - K(u)\|_{L_2}$ to zero is of the form $(cte\rho(\varepsilon))$. This optimal speed of convergence is reached if and only if one selects $N(\varepsilon)\Delta(\varepsilon) \geq cte/\rho^2(\varepsilon)$.

The $(N + k)$ subsampled observations $Y_{n\Delta(\varepsilon)}^\varepsilon$ used to estimate $K(u)$ cover a global observation time span $Span(\varepsilon) = (N + k)\Delta$. For the preceding optimal subsampling schemes, we have $\lim_{\varepsilon\to 0}(Span/N\Delta) = 1$. The minimal observation time span necessary and suffi-

cient to achieve optimal speed of convergence is given by $Span(\varepsilon) = cte/\rho^2(\varepsilon)$. *For each covariance lag* $u > 0$, *this optimal subsampling scheme is realized as follows: first select* $N(\varepsilon)$ *such that* $N(\varepsilon)\rho^2(\varepsilon) \to +\infty$, *then, define covariance lags* $k(\varepsilon)$ *as the closest integers to* $cte \times N(\varepsilon)\rho^2(\varepsilon)$, *and finally, set the subsampling time interval* $\Delta(\varepsilon) = u/k(\varepsilon)$.

Proof. By theorem 2, the difference between approximate and true empirical covariance estimators K_Y^ε and K_X^ε is bounded in L_2-norm by

$$\|K_Y^\varepsilon - K_X^\varepsilon\|_2 \leq 2d\rho(\varepsilon).$$

Applying, to the Gaussian process X, the key result (6.17), we have

$$\|K_X^\varepsilon - K(u)\|_{L_2} \ \leq \ \text{cte}/\sqrt{N(\varepsilon)\Delta(\varepsilon)}, \qquad \forall \varepsilon < 1/cte,$$

where "cte" denotes a positive constant depending only on J, and on the exponential decay coefficients c, a associated to the covariance function K of X. The two last equations imply,

$$\|K_Y^\varepsilon - K(u)\|_{L_2} \ \leq \ 2d\rho(\varepsilon) + \text{cte}/\sqrt{N(\varepsilon)\Delta(\varepsilon)}, \qquad \forall \varepsilon < 1/cte.$$

Since $\rho(\varepsilon)$ is given, the best value achievable for this last upper bound is clearly proportional to $\rho(\varepsilon)$. This optimal upper bound for the L_2-speed of convergence of K_Y^ε to $K(u)$ is reached as soon as $N(\varepsilon)\Delta(\varepsilon) > (\text{cte}/\rho^2(\varepsilon))$.

We show that the L_2-speed of convergence of the form $(\text{cte} \times \rho(\varepsilon))$ cannot be improved in the general Indirect Observability context. To show this, it is sufficient to exhibit at least one specific example of processes Y^ε and X, verifying hypotheses given in section 6.2, such that $\|K_Y^\varepsilon - K(u)\|_{L_2}$ is precisely equivalent to $(\text{cte} \times \rho(\varepsilon))$, as $\varepsilon \to 0$.

The example is constructed as follows. Consider a 2-dimensional centered stationary

Gaussian process $(X_t, Z_t) \in \mathbb{R}^2$ with exponentially decaying covariances,

$$K(u) = E(X_t X_{t+u}), \quad L(u) = E(Z_t Z_{t+u}), \quad M(u) = E(Z_t X_{t+u}).$$

Since X_t, Z_t are assumed to be jointly Gaussian, for each $0 < \varepsilon < 1$, define the centered stationary Gaussian process $Y_t^\varepsilon = X_t + \varepsilon Z_t$. Clearly, the L_4-norms of the processes Y_t^ε, X_t, and Z_t, are bounded. Moreover,

$$\rho(\varepsilon) = \|Y_t^\varepsilon - X_t\|_{L_4} = \varepsilon \|Z_t\|_{L_4} = 3^{1/4} L(0)^{1/2} \varepsilon.$$

Hence, all the hypotheses 6.2 are satisfied by the processes Y_t^ε and X_t. Select $N = N(\varepsilon)$, $k = k(\varepsilon)$ verifying (6.6), (6.7), and set $\Delta = \Delta(\varepsilon) = u/k(\varepsilon)$. By definition of the subsampled covariance estimators K_Y^ε (6.4), we have

$$K_Y^\varepsilon = \frac{1}{N} \sum_{n=0}^{N-1} (X_{n\Delta} + \varepsilon Z_{n\Delta})(X_{n\Delta+u} + \varepsilon Z_{n\Delta+u}) = K_X^\varepsilon + a\varepsilon + b\varepsilon^2,$$

where

$$\begin{aligned} a &= \frac{1}{N} \sum_{n=0}^{N-1} [Z_{n\Delta} X_{n\Delta+u} + X_{n\Delta} Z_{n\Delta+u}], \\ b &= \frac{1}{N} \sum_{n=0}^{N-1} [Z_{n\Delta} Z_{n\Delta+u}], \\ \text{and, therefore, } E(b^2) &= \frac{1}{N^2} \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} E[Z_{m\Delta} Z_{m\Delta+u} Z_{n\Delta} Z_{n\Delta+u}]. \end{aligned}$$

For Gaussian processes, fourth-order moments are expressible in terms of the sum of second-

order moments as follows,

$$E[Z_{m\Delta}Z_{m\Delta+u}Z_{n\Delta}Z_{n\Delta+u}] = L^2(u) + f(n-m),$$

where for each integers j the function $f(j)$ is defined by

$$f(j) = L^2(j\Delta) + L^2(j\Delta + u).$$

Applying the algebraic identity (6.12)

$$(1/N^2)\sum_{m=1}^{N}\sum_{n=1}^{N} f(n-m) = (1/N)\left[\sum_{j=-(N-1)}^{N-1} f(j)\right] - (1/N^2)\left[\sum_{j=1}^{N-1} j(f(j)+f(-j))\right],$$

we obtain

$$E(b^2) = M^2(u) + (1/N)\left[\sum_{j=-(N-1)}^{N-1} f(j)\right] - (1/N^2)\left[\sum_{j=1}^{N-1} j(f(j)+f(-j))\right].$$

The exponential decay assumptions on the covariances K, L, M ensure $|f(j)| < \exp(-\text{cte}|j|\Delta)$. Due to the conditions (6.6), (6.7) on $N(\varepsilon)$, $\Delta(\varepsilon)$, we conclude, just as in section 6.6, that $E(b^2)$ remains uniformly bounded as $\varepsilon \to 0$.

Similar argument shows that $E(a^2)$ is also uniformly bounded as $\varepsilon \to 0$. Moreover, analogous computations prove that as $\varepsilon \to 0$ the coefficient a converges in L_2 to the limit $(M(u)+M(-u))$. Thus, in the limit of $\varepsilon \to 0$, we have the precise asymptotic equivalences,

$$\|K_Y^\varepsilon - K_X^\varepsilon\|_{L_2} = \varepsilon \times (M(u) + M(-u)) + O(\varepsilon^2) \approx \text{cte}\ \rho(\varepsilon).$$

On the other hand, the result (6.15) shows that, as $\varepsilon \to 0$,

$$\|K_X^\varepsilon - K(u)\|_{L_2} \approx \text{cte}\ /\sqrt{N(\varepsilon)\Delta(\varepsilon)}.$$

Therefore, for this class of examples the precise L_2-speeds of convergence are as follows.

1. $N(\varepsilon)\Delta(\varepsilon) \approx \text{cte}/\rho(\varepsilon)^2 \implies \|K_Y^\varepsilon - K(u)\|_{L_2} \approx \text{cte } \rho(\varepsilon)$,

2. $N(\varepsilon)\Delta(\varepsilon) \gg \text{cte}/\rho(\varepsilon)^2 \implies \|K_Y^\varepsilon - K(u)\|_{L_2} \approx \text{cte } \rho(\varepsilon)$, and

3. $N(\varepsilon)\Delta(\varepsilon) \ll \text{cte}/\rho(\varepsilon)^2 \implies \|K_Y^\varepsilon - K(u)\|_{L_2} \gg \text{cte } \rho(\varepsilon)$.

Under the conditions (6.6), (6.7) on $N(\varepsilon)$, $\Delta(\varepsilon)$, the global observation time span, $Span = (N+k)\Delta$, satisfies the limit $(Span/N\Delta) \to 1$, as $\varepsilon \to 0$. Indeed, the validity of (6.23) proved above ensures that the minimal observation time span $Span = cte/\rho^2(\varepsilon)$ is sufficient to achieve optimal speed of convergence. Moreover, condition (3) in the above paragraph establishes the necessity of the result. The remaining concluding statements of the theorem are easily deduced. ■

6.8 Conclusions on Efficient Subsampling Schemes for Indirect Observability

In the Indirect Observability context described in section 6.2, we have shown in section 6.7 that, provided the adaptive subsampling schemes $N(\varepsilon)$, $\Delta(\varepsilon)$ verify (6.2), and $k(\varepsilon)\Delta(\varepsilon)$ is identically equal to u, as $\varepsilon \to 0$; empirical covariances K_Y^ε based on indirect data Y_t^ε are, indeed, L_2-consistent estimators of true covariances $K(u)$.

The optimal L_2-speed of convergence, proportional to $\rho(\varepsilon)$, of these covariance estimators is achieved when $N(\varepsilon)\Delta(\varepsilon) = (\text{cte}/\rho^2(\varepsilon))$. The optimal global observation time span $Span(\varepsilon) = (N+k)\Delta$ for the process Y_t^ε is then proportional to $1/\rho^2(\varepsilon)$. For each $u > 0$, the optimal speed of convergence is inversely proportional to u, since $\Delta(\varepsilon)$ is proportional to u. Moreover, in this case the subsampled covariance estimators K_Y^ε involve *non-vanishing* integer valued covariance lags $k(\varepsilon) \to \infty$.

Based on the results obtained so far, we derive the main corollary concerning the estimation of parameter $\theta \in R^p$. As pointed out above, an estimator for θ is obtained by considering estimation of p covariances $K(u_j)$ of the process X_t, where $u_1 = 0$ and the remaining distinct time lags $u_2, \ldots, u_p$ are all positive.

Theorem 5 *Consider centered stationary observable processes Y_t^ε with arbitrary probability distributions. We assume that as $\varepsilon \to 0$, the Y_t^ε converge in L_4 to an unobservable stationary Gaussian process $X = \{X_t\}$, and verify all the Indirect Observability context of section 6.2. Let $K(u) = K(u, \theta)$ be the covariance function of X_t parametrized by $\theta \in \Theta \subset \mathbb{R}^p$. To estimate θ, consider the system (6.3) associated to the vector $r = (K(u_j, \theta)), j = 1 \ldots p$ of p unknown covariances of X, where the time lags $u_1 = 0$, and $u_2, \ldots, u_p$ are distinct positive real numbers. As above, this system is assumed to have a unique solution $\theta = F(r)$ for all $\theta \in \Theta$. Consider positive integers $N(\varepsilon), k(\varepsilon)$ verifying, as $\varepsilon \to 0$,*

$$N(\varepsilon) \to \infty, \qquad k(\varepsilon) \to \infty, \qquad \text{and} \quad k(\varepsilon)/N(\varepsilon) \to 0. \tag{6.24}$$

Define the time steps,

$$\Delta_j(\varepsilon) = u_j/k(\varepsilon), \quad \text{for} \quad j = 2, \ldots, p.$$

For $j \geq 2$, each covariance $K(u_j)$ is then estimated by the approximate subsampled empirical covariance $K_Y^\varepsilon(u_j)$ based on the $N(\varepsilon)$ observable data $Y^\varepsilon_{n\Delta_j(\varepsilon)}$. For $j = 1$, the variance $K(u_1) = K(0)$ is estimated by the approximate subsampled empirical variance $K_Y^\varepsilon(0)$ based on the $N = N(\varepsilon)$ observable data $Y^\varepsilon_{n\Delta(\varepsilon)}$, where the time interval $\Delta = \Delta(\varepsilon)$ is chosen such that N, Δ satisfy the adaptive subsampling scheme (6.2). Define the vector $\hat{r}(\varepsilon) = (K_Y^\varepsilon(u_0), \ldots, K_Y^\varepsilon(u_p))$. Then, as $\varepsilon \to 0$, under the conditions (6.24), the estimator $\hat{\theta}_\varepsilon = F(\hat{r}(\varepsilon))$ converges in probability to the true value θ.

Proof. The result of theorem 4 implies that, for each $j = 1, \ldots, p$, the subsampled empirical

covariance $K_Y^\varepsilon(u_j)$ converges in L_2 to $K(u_j)$. This combined with the proposition 6.3.1 proves the required convergence. ■

6.9 OU Process as a Special Case under Indirect Observability Context

In chapter 3 we study a special benchmark case of the *indirect observability context* (see section 6.2), specifically, where X_t is considered to be the stationary Ornstein-Uhlenbeck (OU) process, and the approximating process Y_t^ε is the smoothed OU process obtained by local smoothing of X_t, namely

$$Y_t^\varepsilon = \frac{1}{\varepsilon}\int_{t-\varepsilon}^{t} X_s ds. \tag{6.25}$$

Hence, in this special case, both X and Y^ε are centered stationary Gaussian processes. The dynamics of X is given by

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{6.26}$$

and parametrized by

$$\theta = [\gamma, \sigma] \in \Theta = \mathbb{R}^+ \times \mathbb{R}^+.$$

All the hypotheses of the indirect observability context (section 6.2, 6.2) are clearly satisfied for our benchmark case : the X_t have the same Gaussian distribution with mean 0 and variance $(\sigma^2/2\gamma)$, and the covariance function $K(u)$ of X is given by,

$$K(u) = E[X_t X_{u+t}] = \frac{\sigma^2}{2\gamma}\exp(-\gamma|u|), \tag{6.27}$$

which proves the exponential decay of covariances. We then have $E(X_t^4) = 3K(0)^2$, and hence the L_4-norms of all the variables X_t and Y_t^ε are bounded as $\varepsilon \to 0$. We finally evaluate the L_4 speed of convergence $\rho(\varepsilon)$. Since in this benchmark case $X_t - Y_t^\varepsilon$ is Gaussian, we

have

$$\rho(\varepsilon)^4 = E[(X_t - Y_t^\varepsilon)^4] = 3(E[(X_t - Y_t^\varepsilon)^2])^2. \tag{6.28}$$

From (6.25), we derive

$$\begin{aligned} E[(X_t - Y_t^\varepsilon)^2] &= \frac{1}{\varepsilon^2} E\left[\left(\int_{t-\varepsilon}^t (X_t - X_s)ds\right)^2\right] \\ &= \frac{1}{\varepsilon^2}\int_{t-\varepsilon}^t \int_{t-\varepsilon}^t dhds E[(X_t - X_h)(X_t - X_s)], \\ &= \frac{3}{\varepsilon^2}\int_0^\varepsilon \int_0^\varepsilon dhds[K(0) - K(h) - K(s) + K(h-s)]. \end{aligned}$$

Using the explicit expression (6.27) of covariances, this double integral easily yields

$$E[(X_t - Y_t^\varepsilon)^2] = \frac{\sigma^2}{2\gamma}\left[1 - 2\frac{1}{\gamma\varepsilon}(1 - e^{\gamma\varepsilon}) + \frac{1}{\gamma^2\varepsilon^2}(1 - e^{-\gamma\varepsilon})(e^{\gamma\varepsilon} - 1)\right] \approx \frac{\sigma^2}{2}\varepsilon,$$

so that in view of (6.28),

$$\rho(\varepsilon) = \|X_t - Y_t^\varepsilon\|_{L_4} \approx (3/4)^{1/4}\sigma\varepsilon^{1/2}. \tag{6.29}$$

Non-vanishing lags estimation of parameters for approximate OU process

Fix the two covariance lags $u_1 = 0, u_2 = u > 0$. By (6.27), the OU parameter vector $\theta = [\gamma, \sigma^2]$ is a function $F(r)$ of $r = [K(0), K(u)]$, explicitly given by

$$\gamma = (1/u)\ln\left(K(0)/K(u)\right), \qquad \sigma^2 = 2\gamma K(0),$$

Select the numbers of observations $N(\varepsilon)$ and integer valued covariance lags $k(\varepsilon)$ such that, as $\varepsilon \to 0$,

$$N(\varepsilon) \to \infty, \qquad k(\varepsilon) \to \infty, \qquad k(\varepsilon)/N(\varepsilon) \to 0. \tag{6.30}$$

Note that we are here in a subsampling scheme involving *non-vanishing* discrete time lags $k(\varepsilon) \to \infty$, where we define the subsampling time interval by

$$\Delta(\varepsilon) = u/k(\varepsilon), \quad \text{for } u > 0. \tag{6.31}$$

In particular, the hypotheses (6.30) and (6.31) are then equivalent to

$$\varepsilon \to 0, \qquad \Delta(\varepsilon) \to 0, \qquad N(\varepsilon)\Delta(\varepsilon) \to +\infty. \tag{6.32}$$

The observable SOU process Y_t^ε generates then subsampled empirical covariance estimators $\hat{K}_Y^\varepsilon(u)$ and $\hat{K}_Y^\varepsilon(0)$ both defined by

$$\hat{K}_Y^\varepsilon(u) = (1/N) \sum_{n=0}^{N-1} Y_{n\Delta}^\varepsilon Y_{(n\Delta+u)}^\varepsilon, \quad \text{for } u > 0, \quad \text{and for } u = 0.$$

The estimator $\hat{\theta} = (\hat{\gamma}_\varepsilon, \hat{\sigma}_\varepsilon^2)$ of θ is then defined by $\hat{\theta} = F(\hat{K}_Y^\varepsilon(0), \hat{K}_Y^\varepsilon(u))$, so that

$$\hat{\gamma}_\varepsilon = (1/u) \ln \left(\hat{K}_Y^\varepsilon(0) / \hat{K}_Y^\varepsilon(u) \right), \quad \text{and} \quad \hat{\sigma}_\varepsilon^2 = 2\hat{\gamma}_\varepsilon \hat{K}_Y^\varepsilon(0). \tag{6.33}$$

We are considering only the subsampling regime defined by conditions (6.30) and (6.31), which taken together are equivalent to (6.32). In this situation, our generic theorem 5 applies to this benchmark purely Gaussian case X and Y^ε, and shows that the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$ converge in probability to the correct values γ, and σ^2 as $\varepsilon \to 0$.

Moreover the same generic theorem implies that the approximate subsampled covariance estimators converge in L_2 to $K(0)$ and $K(u)$ as $\varepsilon \to 0$, at L_2-speeds of convergence bounded by cte $\times\, \rho(\varepsilon) + 1/\sqrt{N(\varepsilon)\Delta(\varepsilon)}$, where we recall that note $\rho(\varepsilon) = \text{cte}\, \varepsilon^{1/2}$.

Also, specifically for this benchmark example, through an exact computation of the L_2-norms we can deduce a higher order of accuracy and establish convergence in L_2 for the estimators $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$, in particular, $\|\hat{\gamma}_\varepsilon - \gamma\|_{L_2} \sim \text{cte}\, \varepsilon + \text{cte}\, (1/\sqrt{N\Delta})$.

Vanishing Lags estimators for a purely Gaussian benchmark case

In chapter 3, in the same Gaussian benchmark case just presented, we had studied other estimators of θ, based on indirect estimators of covariances $K(0)$ and $K(\Delta(\varepsilon))$ computed from process Y^ε by

$$\hat{K}_Y^\varepsilon(0) = (1/N)\sum_{n=0}^{N-1}(Y_{n\Delta}^\varepsilon)^2, \quad \text{and} \quad \hat{K}_Y^\varepsilon(\Delta) = (1/N)\sum_{n=0}^{N-1} Y_{n\Delta}^\varepsilon Y_{(n+1)\Delta}^\varepsilon.$$

The associated parameter estimators were defined in (3.14) by,

$$\hat{\gamma}_\varepsilon = -\left(\frac{1}{\Delta}\right)\ln\left(\frac{\hat{K}_Y^\varepsilon(\Delta)}{\hat{K}_Y^\varepsilon(0)}\right), \qquad \hat{\sigma}_\varepsilon^2 = -\left(\frac{2\hat{K}_Y^\varepsilon(0)}{\Delta}\right)\ln\left(\frac{\hat{K}_Y^\varepsilon(\Delta)}{\hat{K}_Y^\varepsilon(0)}\right). \tag{6.34}$$

These formulas would be formally identical to those of (6.33) if one could legitimately set $u = \Delta(\varepsilon)$ in the preceding *non-vanishing lags* section, and hence define *vanishing lags* estimators . The asymptotic consistency requirements are however quite distinct. Indeed, in chapter 3, we have shown that if the adaptive subsampling scheme $[N(\varepsilon), \Delta(\varepsilon)]$ verifies,

$$\varepsilon \to 0, \quad \Delta(\varepsilon) \to 0, \quad (\Delta(\varepsilon)/\varepsilon) \to \infty, \quad N(\varepsilon)\Delta(\varepsilon) \to \infty, \tag{6.35}$$

then, the *vanishing lags* estimators $\hat{\gamma}_\varepsilon$ and $\hat{\sigma}_\varepsilon^2$ defined in (6.34) converge in L_2 to the underlying parameters γ, σ^2.

Comparison between vanishing lags and non-vanishing lags estimators

The good subsampling schemes (6.30), (6.31), enabling the existence of consistent indirect parameter estimators of θ based on non-vanishing lags, are obviously *more general* than the subsampling schemes (6.35) warranting, for our specific purely Gaussian benchmark, the existence of consistent indirect estimators based on vanishing lags. Considering subsampling schemes verifying the stronger conditions (6.35), we can then actually use either the non-

vanishing lags estimators or the vanishing lags estimators for the benchmark case of the OU and SOU processes X_t, Y_t^ε.

For our specific benchmark Gaussian case, a comparison of L_2-speeds of convergence for the two types of subsampled parameter estimators shows that non-vanishing lags estimators converge faster than vanishing lags estimators. Namely, considering observation time spans $N\Delta \approx 1/\varepsilon^2$, the best L_2-speeds of convergence of $\hat{\gamma}_\varepsilon$, $\hat{\sigma}_\varepsilon^2$, respectively, to γ and σ^2 are proportional to ε for non-vanishing lags estimators, and to (ε/Δ) for vanishing lags estimators.

Due to the specificity of our purely Gaussian benchmark case, further study is still needed to analyze vanishing lags asymptotics in our generic *indirect observability context*, and to then complete an asymptotic comparison between vanishing lags and non-vanishing lags covariance estimators.

6.10 Concluding Remarks and Future Work

We have presented a generic framework to estimate the parameters θ associated to an underlying stationary Gaussian process X_t, based only on indirect observations generated by an *arbitrary stationary approximating process* $Y_t^\varepsilon \to X_t$ in L_4, as $\varepsilon \to 0$. Note that Y_t^ε is not assumed to be Gaussian.

The vector θ of unknown parameters characterizes the exponentially decaying covariance function $K(u,\theta) = E[X_t X_{t+u}]$. Asymptotically consistent estimation of θ is directly related to the estimation of the covariance function $K(u) = K(u,\theta)$ at finitely many, predetermined and fixed lags $u_1, \ldots, u_p$, where p is the dimension of θ. For a fixed lag $u > 0$, the covariance $K(u)$ is estimated, using discrete observations $U_n = Y^\varepsilon_{n\Delta(\varepsilon)}$, by the *approximate* empirical

covariance estimators K_Y^ε defined as

$$K_Y^\varepsilon = \frac{1}{N(\varepsilon)} \sum_{n=0}^{N-1} U_n U_{n+k(\varepsilon)}.$$

The discrete covariance lag $k(\varepsilon) \to \infty$ with $k(\varepsilon)/N(\varepsilon) \to 0$, and the subsampling time step length $\Delta = \Delta(\varepsilon)$ is defined by

$$\Delta(\varepsilon) = (u/k(\varepsilon)) \to 0.$$

Since $k(\varepsilon)\Delta(\varepsilon) = u$, and the discrete lags $k(\varepsilon) \to \infty$, we refer to these estimators as *non-vanishing lags* estimators. We have derived a precise sufficient condition on $N = N(\varepsilon) \to \infty$, $\Delta = \Delta(\varepsilon) \to 0$, namely $N(\varepsilon)\Delta(\varepsilon) \to \infty$, under which the *approximate* empirical covariance estimators K_Y^ε converge in L_2 to the true covariances $K(u)$.

In this generic *indirect observability context*, the best bound we obtain for the L_2-speed of convergence of these subsampled covariance estimators is proportional to $\rho(\varepsilon) = \|Y_t^\varepsilon - X_t\|_{L_4}$, which is assumed to tend to 0 as $\varepsilon \to 0$. This is achieved by choosing the global time span of observations $Span(\varepsilon) \approx N(\varepsilon)\Delta(\varepsilon) \approx 1/\rho(\varepsilon)^2$. Since there is generally a linear cost attached to the number of observations $N(\varepsilon) \to \infty$, it will be advantageous to choose $\Delta(\varepsilon)$ vanishing as slowly as possible to zero. Precisely, for a fixed $Span = N\Delta$, the choice of a larger N and hence of a smaller Δ, does not improve the accuracy of estimation.

Chapter 3 presented similar questions for a specific purely Gaussian benchmark case, where X_t is the unobserved Ornstein-Uhlenbeck (OU) process and Y_t^ε is a locally *smoothed* OU process. In this purely Gaussian benchmark case, we obtained very precise equivalents for the L_2-speed of convergence of the *vanishing lags* estimators (6.34), but under stronger conditions (6.35) than those studied in the general indirect observability framework to enforce the consistency of non-vanishing lags estimators. For this generic non-Gaussian framework, we intend to develop adequate asymptotics for vanishing lags covariance esti-

mators.

In chapter 5, we study the issue of parameters estimation from indirect approximate data in practical situations, where the actual values of ε and of $\rho(\varepsilon)$ are unknown, since the data are generated by highly intensive simulations of multiscale but high-dimensional ODEs. For prototype examples (the Triad and the Burgers-Hopf equation) we outline multiscale procedures deriving optimal subsampling schemes from one single set of data when ε is unknown. We intend to extend this pragmatic approach to our generic indirect observability context.

Bibliography

[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.*, 56:1–46, 2003.

[2] Y. Ait-Sahalia, P. Mykland, and L. Zhang. A tale of two time scales. *Journal of the American Statistical Association*, 100(472):1394–1411, 2005.

[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(2):351, 2005.

[4] T. Alperovich and A. Sopasakis. Modeling highway traffic with stochastic dynamics. *J.Stat.Phys*, 133:1083–1105, 2008.

[5] T. Andersen, T. Bollerslev, and N. Meddahi. Realized volatility forecasting and market microstructure noise. *Journal of Econometrics*, 2010.

[6] L. Arnold. *Stochastic differential equations: theory and applications.* Wiley-Interscience, New York, 1974.

[7] R. Azencott, A. Beri, and I. Timofeyev. Adaptive sub-sampling for parametric estimation of Gaussian diffusions. *Journal of Statistical Physics*, 139(6):1066–1089, 2010.

[8] R. Azencott, A. Beri, and I. Timofeyev. Parametric estimation of stationary gaussian processes under approximate observability. *Stochastic Processes and their Applications*, 2010. Submitted.

[9] R. Azencott, A. Beri, and I. Timofeyev. Sub-sampling and parametric estimation of multiscale dynamics. *SIAM Multiscale Model. Simul.*, 2010. Submitted.

[10] R. Azencott and D. Dacunha-Castelle. *Series of irregular observations: forecasting and model building*. Springer, 1986.

[11] F. Bandi and P. Phillips. Fully nonparametric estimation of scalar diffusion models. *Econometrica*, 71(1):241–283, 2003.

[12] F. Bandi and J. Russell. Separating microstructure noise from volatility. *Journal of Financial Economics*, 79(3):655–692, 2006.

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

[14] A. Bensoussan, J. Lions, and G. Papanicolaou. *Asymptotic analysis for periodic structures*. North Holland, 1978.

[15] J. Berner. Linking nonlinearity and non-Gaussianity of planetary wave behavior by the Fokker–Planck equation. *Journal of the Atmospheric Sciences*, 62:2098–2117, 2005.

[16] D. Bertsekas, W. Hager, and O. Mangasarian. *Nonlinear programming*. Athena Scientific Belmont, MA, 1999.

[17] J. Bishwal. *Parameter estimation in stochastic differential equations*. Springer Verlag, 2008.

[18] L. Breiman. *Probability*. Addison-Wesley, 1968.

[19] P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer, 1987.

[20] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proc. Nat. Acad. Sci. USA*, 97:2968–2973, 2000.

[21] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction with memory. *Physica D*, 166:239–257, 2002.

[22] A. J. Chorin, A. P. Kast, and R. Kupferman. Unresolved computation and optimal prediction. *Comm. Pure Appl. Math.*, 52:1231–1254, 1999.

[23] K. Chung and R. Williams. *Introduction to stochastic integration*. Springer, 1990.

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

[25] D. Crommelin and E. Vanden-Eijnden. Subgrid-scale parameterization with conditional Markov chains. *J. Atmos. Sci.*, 65:2661–2675, 2008.

[26] D. Crommelin and E. Vanden-Eijnden. Diffusion estimation from multiscale data by operator eigenpairs. *SIAM Multiscale Model. Simul.*, 2010. Submitted.

[27] T. DelSole. A fundamental limitation of Markov models. *J. Atmos. Sci.*, 57:2158–2168, 2000.

[28] P. Deuflhard and C. Schütte. Molecular conformation dynamics and computational drug design. In *Applied mathematics entering the 21st century: invited talks from the ICIAM 2003 Congress*, page 91. Society for Industrial Mathematics, 2004.

[29] J. Doob. The Brownian movement and stochastic equations. *Annals of Mathematics*, 43(2):351–369, 1942.

[30] E. Dynkin. *Markov processes*. Springer-Verlag, 1965.

[31] E. Dynkin. *Diffusions, superdiffusions, and partial differential equations*. Amer Mathematical Society, 2002.

[32] W. Feller. *An introduction to probability theory and its applications, Vol. 2*. Wiley, New York, 1971.

[33] J. Fouque, G. Papanicolaou, and K. Sircar. *Derivatives in financial markets with stochastic volatility*. Cambridge Univ Pr, 2000.

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

[35] M. Freidlin and A. Wentzell. Diffusion processes on graphs and the averaging principle. *The Annals of Probability*, 21(4):2215–2245, 1993.

[36] M. Freidlin and A. Wentzell. *Random perturbations of dynamical systems*. 2nd edition, Springer-Verlag, 1998.

[37] C. W. Gardiner. *Handbook of stochastic methods*. Springer-Verlag, Berlin, 1985.

[38] D. Givon, R. Kupferman, and A. Stuart. Extracting macroscopic dynamics: model problems and algorithms. *Nonlinearity*, 17:R55, 2004.

[39] A. Gloter. Efficient estimation of drift parameters in stochastic volatility models. *Finance and Stochastics*, 11(4):495–519, 2007.

[40] L. Graves. *The theory of functions of real variables*. McGraw-Hill New York, 1956.

[41] A. Griffa, K. Owens, L. Piterbarg, and B. Rozovskii. Estimates of turbulence parameters from Lagrangian data using a stochastic particle model. *Journal of Marine Research*, 53(3):371–401, 1995.

[42] S. Heston. A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of financial studies*, 6(2):327–343, 1993.

[43] E. Hille. Representation of one-parameter semi-groups of linear transformations. *Proceedings of the National Academy of Sciences of the United States of America*, 28(5):175, 1942.

[44] M. Hoffmann. Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Processes and their Applications*, 97(1):147–170, 2002.

[45] I. Horenko, E. Dittmer, A. Fischer, and C. Schütte. Automated model reduction for complex systems exhibiting metastability. *SIAM Multiscale Modeling and Simulation*, 5(3):802–827, 2007.

[46] I. Horenko, C. Hartmann, C. Schütte, and F. Noe. Data-based parameter estimation of generalized multidimensional Langevin processes. *Physical Review E*, 76(1):16706, 2007.

[47] I. Horenko and C. Schütte. Likelihood-based estimation of multidimensional Langevin models and its application to biomolecular dynamics. *Mult. Mod. Sim*, 2008.

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

[49] K. Itô and H. McKean. *Diffusion processes and their sample paths*. Springer Verlag, 1996.

[50] F. Jamshidian. An exact bond option formula. *Journal of Finance*, 44(1):205–209, 1989.

[51] G. Jiang and J. Knight. A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model. *Econometric Theory*, 13(05):615–645, 2009.

[52] I. Karatzas and S. Shreve. *Brownian motion and stochastic calculus*. Springer, 1991.

[53] M. Katsoulakis, A. Majda, and A. Sopasakis. Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 1, deterministic closures. *Comm. Math. Sci.*, 2:255–294, 2004.

[54] M. Katsoulakis, A. Majda, and A. Sopasakis. Multiscale couplings in prototype hybrid deterministic/stochastic systems: Part 2, stochastic closures. *Comm. Math. Sci.*, 3:453–478, 2005.

[55] M. Katsoulakis, A. Majda, and A. Sopasakis. Intermittency, metastability and coarse graining for coupled deterministic–stochastic lattice systems. *Nonlinearity*, 19:1021, 2006.

[56] R. Khasminskii. *Stochastic stability of differential equations*. Sijthoff & Noordhoff Alphen aan den Rijn-Germantown, Md., 1980.

[57] R. Khasminskii and G. Yin. On averaging principles: an asymptotic expansion approach. *SIAM Journal on Mathematical Analysis*, 35:1534, 2004.

[58] P. Kloeden and E. Platen. *Numerical solution of stochastic differential equations*. Springer-Verlag, 1992.

[59] U. Krengel and A. Brunel. *Ergodic theorems*. Walter de Gruyter, 1985.

[60] V. Kulkarni and T. Rolski. Fluid model driven by an Ornstein-Uhlenbeck process. *Probability in the Engineering and Informational Sciences*, 8(03):403–417, 2009.

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

[62] A. J. Majda and I. Timofeyev. Statistical mechanics for truncations of the Burgers-Hopf equation: a model for intrinsic stochastic behavior with scaling. *Milan Journal of Mathematics*, 70(1):39–96, 2002.

[63] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. A mathematics framework for stochastic climate models. *Comm. Pure Appl. Math.*, 54:891–974, 2001.

[64] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. A priori tests of a stochastic mode reduction strategy. *Physica D*, 170:206–252, 2002.

[65] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. Systematic strategies for stochastic mode reduction in climate. *J. Atmos. Sci.*, 60(14):1705–1722, 2003.

[66] A. J. Majda, I. Timofeyev, and E. Vanden-Eijnden. Stochastic models for selected slow variables in large deterministic systems. *Nonlinearity*, 19(4):769–794, 2006.

[67] J. Nielsen, H. Madsen, and P. Young. Parameter estimation in stochastic differential equations: an overview. *Annual Reviews in Control*, 24:83–94, 2000.

[68] K. Nimsaila and I. Timofeyev. Markov Chain Stochastic Parametrizations of Essential Variables. *SIAM Multiscale Model. Simul.*, 2010. Submitted.

[69] J. Nocedal and S. Wright. *Numerical optimization*. Springer-Verlag, 1999.

[70] B. Øksendal. *Stochastic differential equations: an introduction with applications*. Springer Verlag, 2003.

[71] G. Papanicolaou, D. Stroock, and S. Varadhan. Martingale approach to some limit theorems. In *Duke Turbulence Conference (Duke Univ., Durham, NC, 1976), Paper*, volume 6.

[72] A. Papavasiliou. Coarse-grained modeling of multiscale diffusions: the p-variation estimates. *Arxiv preprint arXiv:1002.3241*, 2010.

[73] A. Papavasiliou, G. Pavliotis, and A. Stuart. Maximum likelihood drift estimation for multiscale diffusions. *Stochastic Processes and their Applications*, 119(10):3173–3210, 2009.

[74] E. Pardoux and A. Veretennikov. On Poisson equation and diffusion approximation 1. *Annals of Probability*, 29(3):1061–1085, 2001.

[75] E. Pardoux and A. Veretennikov. On Poisson equation and diffusion approximation 2. *Annals of Probability*, 31(3):1166–1192, 2003.

[76] G. Pavliotis and A. Stuart. Parameter estimation for multiscale diffusions. *Journal of Statistical Physics*, 127(4):741–781, 2007.

[77] G. A. Pavliotis and A. M. Stuart. *Multiscale methods: Averaging and homogenization*. Springer, 2008.

[78] A. Pedersen. A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics*, 22(1):55–71, 1995.

[79] P. Protter. *Stochastic integration and differential equations*. Springer Verlag, 2004.

[80] C. Rao. *Linear statistical inference*. John Wiley and Sons, New York, 1973.

[81] H. Risken. *The Fokker-Plank Equation*. Springer Verlag, Berlin, 2nd edition, 1989.

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

[83] J. Timmer. Parameter estimation in nonlinear stochastic differential equations. *Chaos, Solitons & Fractals*, 11(15):2571–2578, 2000.

[84] R. Tsay. *Analysis of financial time series*. Wiley-Interscience, 2005.

[85] G. Uhlenbeck and L. Ornstein. On the theory of the Brownian motion. *Physical Review*, 36(5):823–841, 1930.

[86] E. Vanden-Eijnden. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Comm. Math. Sci*, 1(2):385–391, 2003.

[87] M. Wang and G. Uhlenbeck. On the theory of the Brownian motion II. *Reviews of Modern Physics*, 17(2-3):323–342, 1945.

[88] A. Yaglom. *An introduction to the theory of stationary random functions*. Dover Publications, 2004.

[89] K. Yosida. On the differentiability and the representation of one-parameter semi-group of linear operators. *Journal of the Mathematical Society of Japan*, 1(1):15–21, 1948.

APPENDICES

0.11 Special Classes of Stochastic Processes

Consider a *filtered probability space*, i.e., a *complete probability space* $(\Omega, \mathcal{F}, P)$ together with a *filtration* $\{\mathcal{F}_t\}$ satisfying the *usual hypotheses*. A family of real-valued random variables $\{X_t\}$ defined on $(\Omega, \mathcal{F}, P)$, where the index $t \in I$ could be discrete ($I = \mathbb{N}$) or continuous-time parameter ($I = \mathbb{R}_+$), is called a stochastic process. A stochastic process is said to be *adapted to a filtration* $\mathbb{F}$, if for each $t \in I$, X_t is $\mathcal{F}_t$ measurable (i.e., $\{X_t < x\} \in \mathcal{F}_t$, for all $x \in \mathbb{R}$). We define a few special classes of stochastic processes.

1. A process X_t is said to be *strictly stationary* if the joint distribution of any arbitrary finite subset of random variables $\{X_{t_1}, X_{t_2}, \ldots, X_{t_k}\}$ is equivalent to the distribution of $\{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_k+h}\}$, for every h such that $t_i + h \in I$, $i = 1, \ldots k$, and for every $k \in \mathbb{N}$.

2. A process X_t is called *Markov* if the conditional probability $P[X_t \in A/\mathcal{F}_s] = P[X_t \in A/X_s]$ for all $t > s$ and for every $A \in \mathcal{F}_t$, i.e., the conditional probability distribution of the process is independent of the past, and only depends on the current value.

3. A process X_t is said to be *Gaussian* if the distribution of any arbitrary finite subset of random variables $\{X_{t_1}, X_{t_2}, \ldots, X_{t_k}\}$ admits a multivariate Gaussian density, for every $k \in \mathbb{N}$.

0.12 Brownian Motion

Consider a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathbb{F} = (\mathcal{F}_t)_{t \geq 0}$ satisfying the usual hypotheses. A stochastic process $\{W(t);\ t \geq 0\}$ is said to be a *Brownian motion* (or *Wiener process*) if it satisfies the following conditions,

- $W(0) = 0$ almost surely.
- $W(t)$ has independent increments, i.e., for $0 \leq t_1 < t_2 < \ldots < t_k < \infty$, the random variables $W(t_1)$, $W(t_2) - W(t_1)$,$\ldots, W(t_k) - W(t_{k-1})$ are independent.
- The increments of $W(t)$ are stationary and Gaussian, i.e., $W(t) - W(s) \sim \mathbf{N}(0, t-s)$, for all $t > s$.
- $W(t)$ has a version with continuous trajectories.

We consider the natural filtration of the Brownian motion i.e. $\mathcal{F}_t = \sigma(W(s); 0 \leq s \leq t) \bigvee \mathcal{N}$, where $\mathcal{N}$ contains the $P-$null sets of $\mathcal{F}$. The Brownian motion $W(t)$ is adapted to the filtration $\mathbb{F}$, i.e., $W(t)$ is $\mathcal{F}_t$-measurable for each t.

There are important properties of the Brownian motion that makes it an attractive tool for modeling uncertainty. For instance, Brownian motion satisfies the martingale property, i.e., $E[W(t)|\mathcal{F}_s] = W(s)$, $\forall t \geq s$. A Brownian motion is square-integrable (i.e., $\|W_t\|_{L_2} < \infty$), Markovian, Gaussian and by definition has independent, stationary increments.

0.13 Kolmogorov Backward and Forward Equations

Consider a one-dimensional diffusion process whose dynamics are given by the following (homogeneous) stochastic differential equation,

$$dX(t) = b(X(t))dt + a(X(t))dW(t). \tag{36}$$

The transition probabilities Q_t defined as,

$$P\left[X(t) \in A \mid X(0) = x\right] = \int_A Q_t(x, dy), \qquad \forall\, A \in \mathcal{F}_t,$$

satisfy the following conditions for every $\delta > 0$,

$$\begin{aligned}
\lim_{h\to 0}\int_{|y-x|\geq\delta}\frac{Q_h(x,dy)}{h} &= 0, \\
\lim_{h\to 0}\int_{|y-x|\geq\delta}\frac{(y-x)Q_h(x,dy)}{h} &= b(x), \qquad \text{(instantaneous } \textit{drift}\text{)}, \\
\lim_{h\to 0}\int_{|y-x|\geq\delta}\frac{(y-x)^2Q_h(x,dy)}{h} &= a^2(x), \qquad \text{(instantaneous } \textit{variance}\text{)}.
\end{aligned} \tag{37}$$

Define the following transformation,

$$u(t,x) = \int_{\mathbb{R}} Q_t(x,dy)u_0(y), \tag{38}$$

that changes, for each fixed $t > 0$, the "initial function" $u_0(x)$ to the function $u(t,x)$. The function $u(t,x)$ may also be expressed as,

$$u(t,x) = E[u_0(X(t))|X(0) = x],$$

which is the conditional expectation with respect to probability measure Q_t. Then, under certain regularity conditions the function $u(t,x)$ satisfies the following PDE [32],

$$\frac{\partial}{\partial t}u(t,x) = b(x)\frac{\partial}{\partial x}u(t,x) + \frac{1}{2}a^2(x)\frac{\partial^2}{\partial x^2}u(t,x), \tag{39}$$

with the initial condition $u(0,x) = u_0(x)$. PDE in (39) is known as the *Kolmogorov backward equation*, and it characterizes the diffusion process $X(t)$ given by (36). The second-order elliptic differential operator $\mathcal{L}$, given by

$$\mathcal{L} = b(x)\frac{\partial}{\partial x} + \frac{1}{2}a^2(x)\frac{\partial^2}{\partial x^2},$$

is the infinitesimal generator of the diffusion process $X(t)$ (See details in [32, 30, 31]), and the backward equation may be expressed as,

$$\frac{\partial}{\partial t}u(t,x) = \mathcal{L}u(t,x).$$

The solution of the backward equation (39) can be written in the semi-group notation [43, 89] as,

$$u(t,x) = e^{\mathcal{L}t}u_0(x). \tag{40}$$

The *Kolmogorov forward equation* (also known as the Fokker-Plank equation [81, 37]) may be derived using the general theory of adjoint partial differential equations. Assume for simplicity that the transition probabilities Q_t have probability densities q_t given by a kernel function $q_t(x,y)$. Then, the transformation in (38) is replaced by,

$$v(s,y) = \int_{\mathbb{R}} v_0(x)q_s(x,y)dx, \tag{41}$$

such that under appropriate regularity conditions $v(s,y)$ satisfies the following PDE,

$$\frac{\partial}{\partial s}v(s,y) = \mathcal{L}^*v(s,y), \tag{42}$$

where the adjoint operator $\mathcal{L}^*$ is given by,

$$\mathcal{L}^*v(s,y) = -\frac{\partial\left[b(y)v(s,y)\right]}{\partial y} + \frac{1}{2}\frac{\partial^2\left[a^2(y)v(s,y)\right]}{\partial y^2}. \tag{43}$$

Here v_0 is an arbitrary probability density, and by definition of q_t, the transform v is again a probability density. Using (41) and (42), we can deduce that,

$$\int_{\mathbb{R}} v_0(x)\left[\frac{\partial q_s(x,y)}{\partial s} - \mathcal{L}^*\, q_s(x,y)\right]dx = 0. \tag{44}$$

For (44) to hold true for arbitrary v_0, the expression within the brackets must vanish, therefore, it is concluded that the transition probability density q_t satisfies the forward equation, namely,

$$\frac{\partial q_s(x,y)}{\partial s} - \mathcal{L}^*\, q_s(x,y) = 0, \tag{45}$$

where $\mathcal{L}^*$ is given by (43). Again, the semi-group notation may be used to express the solution of forward equation (45) as,

$$q_s(x,y) = e^{\mathcal{L}^* s} v_0(x).$$

If the diffusion process $X(t)$ is ergodic, then, the invariant density defined as $q_\infty(y) = \lim_{s\to\infty} q_s(x,y)$, is characterized by $[\mathcal{L}^*\, q_\infty(y)] = 0$.

0.14 Ornstein-Uhlenbeck Process

Ornstein-Uhlenbeck process (denoted here as *OU-process*) is defined as the solution for $t \geq 0$ of the linear SDE

$$dX_t = -\gamma X_t dt + \sigma dW_t, \tag{46}$$

where W_t is the standard Brownian motion and the unknown parameters γ, σ are strictly positive. This linear SDE can be solved by multiplying both sides of the equation by the integrating factor given by $(e^{\gamma t})$, which gives

$$d\left(e^{\gamma t} X_t\right) = \sigma e^{\gamma t} dW_t. \tag{47}$$

The solution X_t of SDE (46), obtained by integrating both sides of (47), is an asymptotically stationary Gaussian process given by

$$X_t = X_0 e^{-\gamma t} + \sigma e^{-\gamma t} \int_0^t e^{\gamma s} dW_s. \tag{48}$$

The Itô integral $M_t = \int_0^t e^{\gamma s} dW_s$ is a centered Gaussian random variable with variance given by the Itô isometry,

$$E\left[(M_t)^2\right] = \int_0^t e^{2\gamma s} ds = \left(\frac{e^{2\gamma t} - 1}{2\gamma}\right). \tag{49}$$

Therefore, when $X_0 = x_0 \in \mathbb{R}$, for each $t > 0$, the random variable X_t is Gaussian, namely,

$$X_t \sim \mathbf{N}\left(x_0 e^{-\gamma t}, \frac{\sigma^2}{2\gamma}\left(1 - e^{-2\gamma t}\right)\right). \tag{50}$$

Using (48) it may be deduced that, for $t > s$,

$$E[X_t X_s] = \sigma^2 e^{-\gamma(t+s)}\, E\left[M_t M_s\right] \;+\; x_0^2 e^{-\gamma(t+s)},$$

where $M_t = \int_0^t e^{\gamma u} dW_u$ is a zero mean martingale. Hence, using the Tower property (double expectations) we obtain,

$$E\left[M_t M_s\right] = E\left[E\left[M_t M_s\right] | \mathcal{F}_s\right] = E\left[(M_s)^2\right].$$

Thus, the covariance function $K(s,t)$ of X_t, for $t > s$, is given as

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

For more detailed presentation of the properties of Itô integral see for instance, [23, 79, 70, 6].

Stationary Moments

The Fokker-Plank equation associated to the OU process (46) is given as,

$$\frac{\partial p}{\partial t} = \gamma \frac{\partial (xp)}{\partial x} + \frac{\sigma^2}{2} \frac{\partial^2 p}{\partial x^2}, \tag{51}$$

with initial distribution given by Dirac point mass δ_{x_0}. The invariant density $\rho(x)$ satisfies,

$$\gamma \frac{\partial(x\rho)}{\partial x} + \frac{\sigma^2}{2}\frac{\partial^2 \rho}{\partial x^2} = 0,$$

and is explicitly given by the Gaussian density, for $\gamma > 0$, with mean 0, and variance $a = \sigma^2/2\gamma$,

$$\rho(x) = \frac{1}{\sqrt{2\pi a}} \exp -x^2/2a.$$

Therefore, for $\gamma > 0$, OU process X_t is asymptotically stationary, and converges almost surely, as $t \to \infty$, to an invariant Gaussian distribution $\mathbf{N}\left(0, \left(\sigma^2/2\gamma\right)\right)$. Under the stationary density the covariance function $K(h)$, for $h \in \mathbb{R}$ is given by

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

Also, when $X_0 \sim \mathbb{N}\left(0, \sigma^2/2\gamma\right)$ and is independent of the filtration $\{\mathcal{F}_t\}$, then X_t is a centered *strictly* stationary Gaussian process, which is evident from the expression in (48). Taking expectations on both sides of (48) shows that the mean of X_t will be zero. Similarly, the variance of X_t will be equal to $\left(\sigma^2/2\gamma\right)$. Since, the second-order moments of the Gaussian process X_t are independent of the time t, therefore, X_t will be strictly stationary.

0.15 Homogenization Method Applied to the Additive Triad Model

Consider the additive triad model [64], namely,

$$\begin{aligned} dx_t &= A_1 y_t z_t \frac{dt}{\varepsilon}, \\ dy_t &= A_2 x_t z_t \frac{dt}{\varepsilon} - g_2 y_t \frac{dt}{\varepsilon^2} + s_2 \frac{dW_1(t)}{\varepsilon}, \\ dz_t &= A_3 x_t y_t \frac{dt}{\varepsilon} - g_3 z_t \frac{dt}{\varepsilon^2} + s_3 \frac{dW_2(t)}{\varepsilon}, \end{aligned} \tag{52}$$

where $A_1 + A_2 + A_3 = 0$ (energy conservation condition), g_i, s_i are known nonnegative parameters, W_1, W_2 are independent Brownian motions, and $\varepsilon > 0$ is the scale separation parameter. Here x_t is the "slow" variable and y_t, z_t are the "fast" variables.

Derivation of the reduced equation corresponding to the triad (52) in the limit of infinite scale separation (i.e., $\varepsilon \to 0$) is as follows. The Kolmogorov backward equation associated to (52) for a scalar function $u = u(t, x, y, z)$ is given as,

$$\frac{\partial}{\partial t} u = \frac{1}{\varepsilon^2} \mathcal{L}_0 \, u + \frac{1}{\varepsilon} \mathcal{L}_1 \, u, \tag{53}$$

where the differential operators $\mathcal{L}_i$ are given by,

$$\mathcal{L}_0 = -g_2 y \frac{\partial}{\partial y} + \frac{1}{2} s_2^2 \frac{\partial^2}{\partial y^2} - g_3 z \frac{\partial}{\partial z} + \frac{1}{2} s_3^2 \frac{\partial^2}{\partial z^2}, \tag{54}$$

$$\mathcal{L}_1 = A_1 y z \frac{\partial}{\partial x} + A_2 x z \frac{\partial}{\partial y} + A_3 x y \frac{\partial}{\partial z}. \tag{55}$$

For brevity (see [77]), assume that the initial condition $u(0, x, y, z) = \phi(x)$, where ϕ is a

scalar function independent of y, z. The differential operator $\mathcal{L}_0$ is the infinitesimal generator for the two dimensional OU process, and admits a bivariate Gaussian invariant density q given by

$$q(y,z) = \frac{\sqrt{g_2 g_3}}{\pi\ s_2 s_3} \exp\left(-\frac{g_2}{s_2^2}y^2 - \frac{g_3}{s_3^2}z^2\right),$$

such that $\mathcal{L}_0^*\ q = 0$. Assume that the null space of $\mathcal{L}_0$ is characterized by $\mathcal{L}_0 \mathbb{I}(y,z)$, where $\mathbb{I}(y,z)$ denotes constants in $y,\ z$.

The multiscale expansion of u is considered, $u = u_0 + \varepsilon u_1 + \varepsilon^2 u_2 + \ldots$, and substituted in (53). Then, comparing the coefficients of powers of ε in (53), gives the following relations,

$$\frac{1}{\varepsilon^2} \quad : \quad \mathcal{L}_0 u_0 = 0, \tag{56}$$

$$\frac{1}{\varepsilon} \quad : \quad \mathcal{L}_0 u_1 + \mathcal{L}_1 u_0 = 0, \tag{57}$$

$$\varepsilon^0 \quad : \quad \mathcal{L}_0 u_2 + \mathcal{L}_1 u_1 = \frac{\partial u_0}{\partial t}. \tag{58}$$

The first relation (56) implies that $u_0 = u_0(x)$, and is constant with respect to the fast variables y, z. The second relation (57) gives rise to the following Poisson equation [77, 74, 75],

$$-\mathcal{L}_0\, u_1 = (\mathcal{L}_1 u_0)\,,$$

such that $\int_{\mathbb{R}^2}\ \mathcal{L}_1 u_0\ q(y,z)\, dy\, dz = 0$, therefore, the solution u_1 can be formally expressed as,

$$u_1 = -\mathcal{L}_0^{-1}\ (\mathcal{L}_1 u_0)\,, \tag{59}$$

where the operator $\mathcal{L}_0^{-1}$ is given as,

$$\mathcal{L}_0^{-1} = -\int_0^\infty\ e^{\mathcal{L}_0\, t} dt. \tag{60}$$

Now, let $\mathbb{P}$ denote the expectation with respect to the invariant density $q(y,z)$, then using the Feynman-Kaĉ technique and applying the expectation operator $\mathbb{P}$ to both sides of the relation (58) gives,

$$\mathbb{P}(\mathcal{L}_0 u_2) + \mathbb{P}(\mathcal{L}_1 u_1) = \mathbb{P}\left(\frac{\partial u_0}{\partial t}\right),$$

which, since, $u_0 = u_0(x)$, $\mathcal{L}_0^*\, q = 0$, and using (59); reduces to the following,

$$-\mathbb{P}\left(\mathcal{L}_1 \mathcal{L}_0^{-1}\ (\mathcal{L}_1 u_0)\right) = \frac{\partial u_0}{\partial t}. \tag{61}$$

The right hand side of (61) may be computed as follows,

$$\begin{aligned}
-\mathbb{P}\left(\mathcal{L}_1 \mathcal{L}_0^{-1}\ (\mathcal{L}_1 u_0)\right) &= -\mathbb{P}\left(A_1 \mathcal{L}_1 \left(\frac{\partial u_0}{\partial x}\mathcal{L}_0^{-1}(yz)\right)\right) \\
&= -A_1^2 \frac{\partial^2 u_0}{\partial x^2}\mathbb{P}\left(yz\mathcal{L}_0^{-1}(yz)\right) - A_1 A_2 x \frac{\partial u_0}{\partial x}\mathbb{P}\left(z\frac{\partial\left(\mathcal{L}_0^{-1}(yz)\right)}{\partial y}\right) + \dots \\
&\quad - A_1 A_3 x \frac{\partial u_0}{\partial x}\mathbb{P}\left(y\frac{\partial\left(\mathcal{L}_0^{-1}(yz)\right)}{\partial z}\right).
\end{aligned} \tag{62}$$

Using (40) (60), the term $\mathbb{P}\left(yz\mathcal{L}_0^{-1}(yz)\right)$ is equal to,

$$\begin{aligned}
\mathbb{P}\left(yz\mathcal{L}_0^{-1}(yz)\right) &= -\int_0^\infty \mathbb{P}(y(t)y(0))\mathbb{P}(z(t)z(0))dt \\
&= -\frac{s_2^2 s_3^2}{4g_2 g_3}\int_0^\infty e^{-(g_2+g_3)t}\, dt \\
&= -\frac{s_2^2 s_3^2}{4g_2 g_3(g_2+g_3)}.
\end{aligned} \tag{63}$$

Using the fact that the Gaussian density $q(y,z)$ vanishes exponentially as $|y|, |z| \to \infty$, and

applying integration by parts, we obtain,

$$\begin{aligned}
\mathbb{P}\left(z\frac{\partial\left(\mathcal{L}_0^{-1}(yz)\right)}{\partial y}\right) &= \frac{2g_2}{s_2^2}\mathbb{P}\left(yz\left(\mathcal{L}_0^{-1}(yz)\right)\right) = -\frac{s_3^2}{2g_3(g_2+g_3)},\\
\mathbb{P}\left(y\frac{\partial\left(\mathcal{L}_0^{-1}(yz)\right)}{\partial z}\right) &= \frac{2g_3}{s_3^2}\mathbb{P}\left(yz\left(\mathcal{L}_0^{-1}(yz)\right)\right) = -\frac{s_2^2}{2g_2(g_2+g_3)}.
\end{aligned} \tag{64}$$

Finally, substituting (64), (63) in (62), we obtain the following backward equation verified by u_0,

$$\frac{\partial u_0}{\partial t} = \left[\frac{A_1^2\, s_2^2 s_3^2}{4g_2g_3\,(g_2+g_3)}\right]\frac{\partial^2 u_0}{\partial x^2} + \left[\frac{A_1 x}{2(g_2+g_3)}\left(\frac{A_2 s_3^2}{g_3}+\frac{A_3 s_2^2}{g_2}\right)\right]\frac{\partial u_0}{\partial x},$$

which characterizes the OU process,

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

where the parameters γ, σ are given by,

$$\gamma = -\frac{A_1}{2(g_2+g_3)}\left(\frac{A_2 s_3^2}{g_3}+\frac{A_3 s_2^2}{g_2}\right), \qquad \sigma = \frac{|A_1| s_2 s_3}{\sqrt{2g_2g_3\,(g_2+g_3)}}.$$

The limiting OU process is asymptotically stationary provided $\gamma > 0$, which is a restriction on the choice of parameters in the full triad model. Hence, we finally deduce that the slow variable x_t in the triad weakly converges to the OU process X_t, as $\varepsilon \to 0$. See [77] for a detailed discussion of the method of homogenization for a generic multiscale system.