Maximum Likelihood Estimation for Multiscale Ornstein-Uhlenbeck Processes

Anastasia Papavasiliou *
Department of Statistics
University of Warwick
and
Fan Zhang †
International Monetary Fund

April 10, 2015

Abstract

We study the problem of estimating the parameters of an Ornstein-Uhlenbeck (OU) process that is the coarse-grained limit of a multiscale system of OU processes, given data from the multiscale system. We consider both the averaging and homogenization cases and both drift and diffusion coefficients. By restricting ourselves to the OU system, we are able to substantially improve the results in [26, 23] and provide some intuition of what to expect in the general case.

Keywords : multiscale diffusions, Ornstein-Uhlenbeck process, parameter estimation, maximum likelihood, subsampling.

1 Introduction

A necessary step in statistical modelling is to fit the chosen model to the data by inferring the value of the unknown parameters. In the case of stochastic differential equations (SDE), this is a well studied problem [7, 17, 27]. However, quite often, there is a mismatch between model and data. The actual system the data comes from is often of multiscale nature whilst the SDE we are fitting is only an approximation of its behavior at a certain scale. This phenomenon has been observed in many applications, ranging from econometrics [1, 2, 21] to chemical engineering [5] and molecular dynamics [26]. In this paper, we study how this inconsistency between the coarse-grained model that we fit and the microscopic dynamics from which the data is generated affects the estimation problem.

In this paper, we take the approach by minimizing the discrepancy between the maximum likelihood estimators based on the multiscale and approximated systems, with our focus on the drift and diffusion parameters of both averaging and homogenization. There are existing literatures explored alternatives to achieve a certain part of our goal. [3, 4] explored an approach to estimate the bias between the estimators based

*Email: A.Papavasiliou@warwick.ac.uk
†Email: FZhang@imf.org

Paper No. 15-05, www.warwick.ac.uk/go/crism

the multiscale and approximated OU processes, as a function of the subsampling step size and the scale factor. However, their approach is some what ad-hoc by limited to scalar systems. [18] also discussed on achieving the strong convergence of the estimators of the averaged multiscale OU system, and focused on the strong convergence of the diffusion parameter.

This problem has also been discussed in [26, 23]. Our aim is to strengthen the results in [26, 23]. To achieve this, we only consider the case where the multiscale system is an Ornstein-Uhlenbeck process, where the averaging and homogenization principles still hold. This allows us to prove a stronger mode of convergence for the asymptotics.

To be more specific, we will consider multiscale systems of SDEs of the form

$$\frac{dx}{dt} = a_{11}x + a_{12}y + \sqrt{q_1}\frac{dU}{dt} \tag{1a}$$

$$\frac{dy}{dt} = \frac{1}{\epsilon}\left(a_{21}x + a_{22}y\right) + \sqrt{\frac{q_2}{\epsilon}}\frac{dV}{dt} \tag{1b}$$

or

$$\frac{dx}{dt} = \frac{1}{\epsilon}\left(a_{11}x + a_{12}y\right) + \left(a_{13}x + a_{14}y\right) + \sqrt{q_1}\frac{dU}{dt} \tag{2a}$$

$$\frac{dy}{dt} = \frac{1}{\epsilon^2}\left(a_{21}x + a_{22}y\right) + \sqrt{\frac{q_2}{\epsilon^2}}\frac{dV}{dt} \tag{2b}$$

We refer to equations (1) as the averaging problem, and to equations (2) as the homogenization problem. We assume that in both cases the averaging or homogenization limits exist. In both cases, it will be of the form

$$\frac{dX}{dt} = aX + \sqrt{\sigma}\frac{dW}{dt}, \quad a < 0, \sigma > 0, \tag{3}$$

for appropriate a and σ. Our goal will be to estimate a and σ, assuming that we continuously observe x from (1) or (2). It is a well known result (see [7, 20]) that, given X, the maximum likelihood estimators for a is

$$\hat{a}_T = \left(\int_0^T X dX\right)\left(\int_0^T X^2 dt\right)^{-1}. \tag{4}$$

If X is discretely observed, then the maximum likelihood estimator of σ is

$$\hat{\sigma}_\delta = \frac{1}{T}\sum_{n=0}^{N-1}\left(X_{(n+1)\delta} - X_{n\delta}\right)^2 \tag{5}$$

which converges a.s. to σ as $\delta \to 0$, i.e. if X is observed continuously, then σ will be known. Our approach will be to still use the estimators defined in (4) and (5), replacing X by its x approximation coming from the multiscale model and then studying their asymptotic properties. In section 2, we discuss the averaging case, where the data comes from equation (1a) while in section 3 we study the homogenization case corresponding to equation (2a).

We shall discuss problems in scalars for simplicity of notation and writing. However, the conclusions can easily be extended to finite dimensions. We will use c to denote an arbitrary constant which can vary from occurrence to occurrence. Also, for the sake of simplicity we will sometimes write x_n (or y_n, X_n) instead of $x(n\delta)$ (resp. $y(n\delta)$, $X(n\delta)$). Finally, note that the transpose of an arbitrary matrix A is denoted by A^*.

2 Averaging

We consider the system of stochastic differential equations described by (1) (averaging case), for the variables $(x, y) \in \mathcal{X} \times \mathcal{Y}$. We may take $\mathcal{X}$ and $\mathcal{Y}$ as either $\mathbb{R}$ or $\mathbb{T}$. Our interest is in data generated by the projection onto the x coordinate of the system. We will make the following

Assumptions 2.1. *(i) U, V are independent Brownian motions;*

(ii) q_1, q_2 are positive;

(iii) $0 < \epsilon \ll 1$;

(iv) $a_{22} < 0$ and $a_{11} < a_{12}a_{22}^{-1}a_{21}$;

(v) $x(0)$ and $y(0)$ independent of U and V, $\mathbb{E}\left(x(0)^2 + y(0)^2\right) < \infty$.

In what follows, we will refer to the following equation as the *averaged equation* for system (1):

$$\frac{dX}{dt} = \tilde{a}X + \sqrt{q_1}\frac{dU}{dt} \tag{6}$$

where:

$$\tilde{a} = a_{11} - a_{12}a_{22}^{-1}a_{21} \tag{7}$$

2.1 The Paths

In this section, we show that the projection of system (1) onto the x coordinate converges in a strong sense to the solution X of the averaged equation (6). Our result extends that of [25] (Theorem 17.1) where the state space $\mathcal{X}$ is restricted to $\mathbb{T}$ and the averaged equation is deterministic. Assuming that the system is an OU process, the domain can be extended to $\mathbb{R}$ and the averaged equation can be stochastic. We prove the following lemma first:

Lemma 2.2. *Suppose that (x, y) solves (1a) and Assumptions 2.1 are satisfied. Then, for finite $T > 0$ and ϵ small,*

$$\mathbb{E} \sup_{0\le t\le T} \left(x(t)^2 + y(t)^2\right) \approx \mathcal{O}\left(\log\left(1 + \frac{T}{\epsilon}\right)\right). \tag{8}$$

Proof. Since U and V are independent, we can rewrite (1) in vector form as

$$d\mathbf{x}_t = \mathbf{a}\mathbf{x}_t dt + \sqrt{\mathbf{q}}d\mathbf{W}_t \tag{9}$$

where

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \mathbf{a} = \begin{pmatrix} a_{11} & a_{12} \\ \frac{1}{\epsilon}a_{21} & \frac{1}{\epsilon}a_{22} \end{pmatrix}, \mathbf{q} = \begin{pmatrix} q_1 & 0 \\ 0 & \frac{q_2}{\epsilon} \end{pmatrix}$$

and $\mathbf{W} = (U, V)$ is two-dimensional Brownian motion. Given the form of $\mathbf{a}$, it is an easy exercise to show that its eigenvalues will be of order $\mathcal{O}(1)$ and $\mathcal{O}(\frac{1}{\epsilon})$. Therefore, we define the eigenvalue decomposition of $\mathbf{a}$ as

$$\mathbf{a} = PDP^{-1} \text{ with } D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \frac{1}{\epsilon}\lambda_2 \end{pmatrix}.$$

Again, it is not hard to see that if (p_1, p_2) is an eigenvector, $\mathcal{O}(p_1) = \mathcal{O}(\lambda^{-1}p_2)$. So, for the eigenvector corresponding to eigenvalue of order $\mathcal{O}(1)$, all elements of the eigenvector will also be of order $\mathcal{O}(1)$ while for the eigenvector corresponding to eigenvalue of order $\mathcal{O}(1/\epsilon)$, we will have that $p_1 \sim \mathcal{O}(1)$ and $p_2 \sim \mathcal{O}(\epsilon)$.

Now, let us define $\Sigma = P^{-1}\mathbf{q}(P^{-1})^*$. It follows that

$$\Sigma = \begin{pmatrix} \mathcal{O}(1) & \mathcal{O}(1) \\ \mathcal{O}(1) & \mathcal{O}(1/\epsilon) \end{pmatrix}$$

We apply a linear transformation to the system of equations (9) so that the drift matrix becomes diagonal. It follows form [12] that

$$\mathbb{E}\left(\sup_{0\le t\le T} \|\mathbf{x}(t)\|^2\right) \le C\frac{\log\left(1+\max_i(|D_{ii}|)T\right)}{\min_i(|D_{ii}/\Sigma_{ii}|)}, i \in \{1,2\}.$$

Since the diagonal elements of D and Σ are of the same order and $\max_i |D_{ii}| = \mathcal{O}(\frac{1}{\epsilon})$, we have

$$\mathbb{E}\left(\sup_{0\le t\le T} \|\mathbf{x}(t)\|^2\right) = \mathcal{O}\left(\log(1+T/\epsilon)\right).$$

Finally, since $\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}$, we get

$$\mathbb{E}\left(\sup_{0\le t\le T}\left(\|x(t)\|^2 + \|y(t)\|^2\right)\right) = \mathcal{O}\left(\log(1+\frac{T}{\epsilon})\right) .$$

This completes the proof. □

Theorem 2.3. *Let Assumptions 2.1 hold for system* (1)*. Suppose that x and X are two solutions of (1a) and* (6) *respectively, corresponding to the same realization of the U process and $x(0) = X(0)$. Then, x converges to X in $L^2(\Omega, C([0,T], \mathcal{X}))$. More specifically,*

$$\mathbb{E}\sup_{0\le t\le T}(x(t) - X(t))^2 \le c\left(\epsilon^2 log\left(\frac{T}{\epsilon}\right) + \epsilon T\right)e^T ,$$

when T is fixed finite, the above bound can be simplified to

$$\mathbb{E}\sup_{0\le t\le T}(x(t) - X(t))^2 = \mathcal{O}(\epsilon) .$$

Proof. For auxiliary equations used in the proof, please refer to the construction in [25]. The generator of system (1) is

$$\mathcal{L}_{avg} = \frac{1}{\epsilon}\mathcal{L}_0 + \mathcal{L}_1,$$

where

$$\begin{aligned}\mathcal{L}_0 &= (a_{21}x + a_{22}y)\frac{\partial}{\partial y} + \frac{1}{2}q_2\frac{\partial^2}{\partial y^2}\\ \mathcal{L}_1 &= (a_{11}x + a_{12}y)\frac{\partial}{\partial x} + \frac{1}{2}q_1\frac{\partial^2}{\partial x^2}\end{aligned}$$

To prove that the L^2 error between the solutions $x(t)$ and $X(t)$ is of order $\mathcal{O}(\sqrt{\epsilon})$, we first need to find the function $\Phi(x, y)$ which solves the Poisson equation

$$-\mathcal{L}_0\Phi = a_{11}x + a_{12}y - \tilde{a}x\ ,\ \int_{\mathcal{Y}}\Phi\rho(y;x)dy = 0; \tag{10}$$

where $\rho(y; x)$ is the invariant density of y in (1b) with x fixed. In this case, the partial differential equation (10) is linear and can be solved explicitly

$$\Phi(x, y) = \Phi(y) = -(a_{12}a_{22}^{-1})y. \tag{11}$$

Applying Itô formula to $\Phi(x, y)$, we get

$$\frac{d\Phi}{dt} = \frac{1}{\epsilon}\mathcal{L}_0\Phi + \mathcal{L}_1\Phi + \frac{1}{\sqrt{\epsilon}}\sqrt{q_2}\frac{\partial}{\partial y}\Phi\frac{dV_t}{dt},$$

and substituting into (1a) gives

$$\begin{aligned}\frac{dx}{dt} &= (\tilde{a}x - \mathcal{L}_0\Phi) + \sqrt{q_1}\frac{dU_t}{dt}\\ &= \tilde{a}x - \epsilon\frac{d\Phi}{dt} + \epsilon\mathcal{L}_1\Phi + \sqrt{\epsilon}\sqrt{q_2}\frac{\partial}{\partial y}\Phi\frac{dV_t}{dt} + \sqrt{q_1}\frac{dU_t}{dt}.\end{aligned} \tag{12}$$

Define

$$\theta(t) := (\Phi(x(t), y(t)) - \Phi(x(0), y(0))) - \int_0^t (a_{11}x(s) + a_{12}y(s))\frac{\partial\Phi}{\partial x}ds.$$

From (11), we see that Φ does not depend on x and thus

$$\begin{aligned}\theta(t) &= \Phi(x(t), y(t)) - \Phi(x(0), y(0))\\ &= -(a_{12}a_{22}^{-1})(y(t) - y(0)).\end{aligned} \tag{13}$$

Now define

$$\begin{aligned}M(t) &:= -\int_0^t \sqrt{q_2}\frac{\partial}{\partial y}\Phi(x(s), y(s))dV_s\\ &= -\int_0^t \sqrt{q_2}(a_{12}a_{22}^{-1})dV_s.\end{aligned}$$

Itô isometry gives

$$\mathbb{E}M^2(t) = ct \tag{14}$$

The solution of (1a) in the form of (12) is

$$x(t) = x(0) + \int_0^t \tilde{a}x(s)ds + \epsilon\theta(t) + \sqrt{\epsilon}M(t) + \sqrt{q_1}\int_0^t dU_s\ .$$

Also, from the averaged equation (6), we get

$$X(t) = X(0) + \int_0^t \tilde{a} X(s) ds + \sqrt{q_1} \int_0^t dU_s \, .$$

Let $e(t) = x(t) - X(t)$. By assumption, $e(0) = 0$ and

$$e(t) = \int_0^t \tilde{a} \left(x(s) - X(s) \right) ds + \epsilon \theta(t) + \sqrt{\epsilon} M(t) \, . \tag{15}$$

Then,

$$e^2(t) \leq 3 \left(\tilde{a} \int_0^t e(s) ds \right)^2 + 3\epsilon^2 \theta^2(t) + 3\epsilon M^2(t) \, .$$

Apply Lemma 2.2 on (15), the Burkholder-Davis-Gundy inequality [25] and Hölder inequality, we get

$$\begin{aligned} & \mathbb{E} \left(\sup_{0 \leq t \leq T} e^2(t) \right) \\ \leq \quad & c \left(\int_0^T \mathbb{E} e^2(s) ds + \epsilon^2 \log(\frac{T}{\epsilon}) + \epsilon T \right) \\ \leq \quad & c \left(\epsilon^2 \log(\frac{T}{\epsilon}) + \epsilon T + \int_0^T \mathbb{E} \sup_{0 \leq u \leq s} e^2(u) ds \right) . \end{aligned}$$

By Gronwall's inequality [25], we deduce that

$$\mathbb{E} \left(\sup_{0 \leq t \leq T} (e(t))^2 \right) \leq c(\epsilon^2 \log(\frac{T}{\epsilon}) + \epsilon T) e^T .$$

When T is fixed, we have

$$\mathbb{E} \left(\sup_{0 \leq t \leq T} (e(t))^2 \right) = \mathcal{O}(\epsilon) \, .$$

This completes the proof. □

2.2 The Drift Estimator

Suppose that we want to estimate the drift of the process X described by (6) but we only observe a solution $\{x(t)\}_{t \in (0,T)}$ of (1a). According to the previous theorem, x is a good approximation of X, so we replace X in the formula of the MLE (4) by x. In the following theorem, we show that the error we will be making is insignificant, in a sense to be made precise.

Theorem 2.4. *Suppose that x is the projection to the x-coordinate of a solution of system* (1) *satisfying Assumptions 2.1. Let $\hat{a}_T^\epsilon$ be the estimate we get by replacing X in* (4) *by x, i.e.*

$$\hat{a}_T^\epsilon = \left(\int_0^T x dx \right) \left(\int_0^T x^2 dt \right)^{-1} . \tag{16}$$

Then,

$$\lim_{\epsilon \to 0} \lim_{T \to \infty} \mathbb{E} (\hat{a}_T^\epsilon - \tilde{a})^2 = 0 \, .$$

Proof. We define

$$I_1 = \frac{1}{T}\int_0^T x dx \;\text{ and }\; I_2 = \frac{1}{T}\int_0^T x^2 dt.$$

By ergodicity, which is guaranteed by Assumptions 2.1 *(iii) and (iv)*

$$\lim_{T\to\infty} I_2 = \mathbb{E}(x^2) = C \neq 0 \; a.s.,$$

which is a non-zero constant. We expand dx using Itô formula [25] applied on Φ as in (12):

$$I_1 = J_1 + J_2 + J_3 + J_4 + J_5$$

where

$$J_1 = \frac{1}{T}\int_0^T \tilde{a} x^2 dt$$

$$J_2 = \frac{\epsilon}{T}\int_0^T x d\Phi$$

$$J_3 = \frac{\epsilon}{T}\int_0^T \mathcal{L}_1 \Phi x dt$$

$$J_4 = \frac{\sqrt{\epsilon}}{T}\int_0^T \frac{\partial}{\partial y}\Phi \sqrt{q_2} x dV_t$$

$$J_5 = \frac{1}{T}\sqrt{q_1}\int_0^T x dU_t$$

It is obvious that

$$J_1 = \tilde{a} I_2.$$

Since Φ is linear in y, and by Itô isometry, we get

$$\begin{aligned} \mathbb{E}\left(J_4^2\right) &= \frac{c\epsilon}{T}\mathbb{E}\left(\frac{1}{T}\int_0^T x(t) dV_t\right)^2 \\ &= \frac{c\epsilon}{T}\mathbb{E}\left(\frac{1}{T}\int_0^T x^2(t) dt\right) \end{aligned}$$

by ergodicity, we have

$$\mathbb{E}\left(J_4^2\right) = \frac{c\epsilon}{T}.$$

Similarly for J_5,

$$\begin{aligned} \mathbb{E}\left(J_5^2\right) &= \frac{c}{T}\mathbb{E}\left(\frac{1}{T}\int_0^T x(t) dU_t\right)^2 \\ &= \frac{c}{T} \end{aligned}$$

We know Φ is independent of x, so

$$J_3 \equiv 0.$$

Finally, using (11) and (1b) we break J_2 further into

$$J_2 = -\frac{1}{T}\int_0^T (a_{12}a_{22}^{-1})(a_{21}x + a_{22}y)xdt - \frac{a_{12}a_{22}^{-1}\sqrt{\epsilon q_2}}{T}\int_0^T xdV_t$$

Again, using Itô isometry and ergodicity, we bound the L^2 norm of the second term by

$$\mathbb{E}\left(\frac{a_{12}a_{22}^{-1}\sqrt{\epsilon q_2}}{T}\int_0^T xdV_t\right)^2 \leq \frac{c\epsilon}{T}.$$

By ergodicity, the first term converges in L^2 as $T \to \infty$,

$$-\frac{a_{12}a_{22}^{-1}}{T}\int_0^T (a_{21}x + a_{22}y)xdt \to -a_{12}\mathbb{E}_{\rho^\epsilon}\left((a_{22}^{-1}a_{21}x + y)x\right).$$

We write the expectation as

$$\mathbb{E}_{\rho^\epsilon}\left((a_{22}^{-1}a_{21}x + y)x\right) = \mathbb{E}_{\rho^\epsilon}\left(\mathbb{E}_{\rho^\epsilon}\left((a_{22}^{-1}a_{21}x + y)x|x\right)\right)$$

Clearly, the limit of ρ^ϵ conditioned on x is a normal distribution with mean $-a_{22}^{-1}a_{21}x$. Thus, we see that

$$\lim_{\epsilon\to 0}\mathbb{E}_{\rho^\epsilon}\left((a_{22}^{-1}a_{21}x + y)x\right) = 0.$$

Putting everything together, we see that

$$\lim_{\epsilon\to 0}\lim_{T\to\infty}(I_1 - \tilde{a}I_2) = 0 \quad \text{in} \quad L^2$$

Since the denominator I_2 of $\hat{a}_T^\epsilon$ converges almost surely, the result follows. □

2.3 Asymptotic Normality for the Drift Estimator

We extend the proof of Theorem 2.4 to prove asymptotic normality for the estimator $\hat{a}_T^\epsilon$. We have seen that

$$\hat{a}_T^\epsilon - \tilde{a} = \frac{J_2 + J_4 + J_5}{I_2}.$$

We will show that

$$\sqrt{T}\left(\hat{a}_T^\epsilon - \tilde{a} + a_{12}\mathbb{E}_{\rho^\epsilon}\left(x(a_{22}^{-1}a_{21}x + y)\right)\right) \to \mathcal{N}\left(0, \sigma_\epsilon^2\right)$$

and compute the limit of σ_ϵ^2 as $\epsilon \to 0$. First we apply the Central Limit Theorem for martingales to J_4 and J_5 (see [13]). We find that

$$\sqrt{T}J_4 \to \mathcal{N}\left(0, \sigma(4)_\epsilon^2\right) \quad \text{as } T \to \infty$$

where

$$\sigma(4)_\epsilon^2 = \epsilon q_2(a_{12}a_{22}^{-1})^2\mathbb{E}_{\rho^\epsilon}x^2$$

and

$$\sqrt{T}J_5 \to \mathcal{N}\left(0, \sigma(5)_\epsilon^2\right) \quad \text{as } T \to \infty$$

where

$$\sigma(5)_\epsilon^2 = q_1\mathbb{E}_{\rho^\epsilon}x^2.$$

We write $J_2 = J_{2,1} + J_{2,2}$ where

$$J_{2,1} = -\frac{a_{12}a_{22}^{-1}}{T}\int_0^T (a_{21}x^2 + a_{22}xy)dt \text{ and } J_{2,2} = -\frac{a_{12}a_{22}^{-1}\sqrt{\epsilon q_2}}{T}\int_0^T xdV.$$

Once again, we apply the Central Limit Theorem for martingales to $J_{2,2}$ and we find

$$\sqrt{T}J_{2,2} \to \mathcal{N}\left(0, \sigma(2,2)_\epsilon^2\right) \text{ as } T \to \infty$$

where

$$\sigma(2,2)_\epsilon^2 = \epsilon(a_{21}a_{22}^{-1})^2 q_2 \mathbb{E}_{\rho^\epsilon} x^2.$$

Finally, we apply the Central Limit Theorem for functionals of ergodic Markov Chains to $J_{2,1}$ (see [8]). We get

$$\sqrt{T}\left(J_{2,1} + a_{12}\mathbb{E}_{\rho^\epsilon}\left(x(a_{22}^{-1}a_{21}x + y)\right)\right) \to \mathcal{N}\left(0, \sigma(2,1)_\epsilon^2\right)$$

as $T \to \infty$, where

$$\sigma(2,1)_\epsilon^2 = \int_{\mathcal{X}\times\mathcal{Y}} \xi^2(x,y)\rho_\epsilon(x,y)dxdy + 2\int_{\mathcal{X}\times\mathcal{Y}} \xi(x,y)\int_0^\infty (P_t^\epsilon \xi)(x,y)dt\rho_\epsilon(x,y)dxdy$$

with

$$\xi(x,y) = -\left(a_{12}a_{22}^{-1}a_{21}x^2 + a_{12}xy\right) + \mathbb{E}\left(a_{12}a_{22}^{-1}a_{21}x^2 + a_{12}xy\right)$$

and

$$(P_t^\epsilon \xi)(x,y) = \mathbb{E}\left(\xi(x(t), y(t)) | x(0) = x, y(0) = y\right).$$

Putting everything together, we get that as $T \to \infty$,

$$\sqrt{T}(J_2 + J_4 + J_5) \to X_{2,1} + X_{2,2} + X_4 + X_5$$

in law, where $X_i \sim \mathcal{N}(0, \sigma(i)_\epsilon^2)$ for $i \in \{\{2,1\}, \{2,2\}, 4, 5\}$. Finally, we note that the denominator I_2 converges almost surely as $T \to \infty$ to $\mathbb{E}_{\rho^\epsilon}(x(t)^2)$. It follows from Slutsky's theorem that as $T \to \infty$,

$$\sqrt{T}\left(\hat{a}_T^\epsilon - \tilde{a} + a_{12}\mathbb{E}_{\rho^\epsilon}\left(x(a_{22}^{-1}a_{21}x + y)\right)\right) \to X_\epsilon$$

in law, where

$$X_\epsilon = \frac{X_{2,1} + X_{2,2} + X_4 + X_5}{\mathbb{E}_{\rho^\epsilon}(x(t)^2)} \sim \mathcal{N}(0, \sigma_\epsilon^2).$$

It remains to compute $\lim_{\epsilon\to 0}\sigma_\epsilon^2$. We have already seen that $\sigma(2,2)_\epsilon^2 \sim \mathcal{O}(\epsilon)$ and $\sigma(4)_\epsilon^2 \sim \mathcal{O}(\epsilon)$. Thus, we need to compute

$$\lim_{\epsilon\to 0}\mathbb{E}(X_{2,1} + X_5)^2 = \lim_{\epsilon\to 0}\mathbb{E}\left(X_{2,1}^2 + 2X_{2,1}X_5 + X_5^2\right).$$

First, we see that

$$\lim_{\epsilon\to 0}\mathbb{E}(X_5^2) = q_1 \lim_{\epsilon\to 0}\mathbb{E}_{\rho_\epsilon} x^2 = q_1\mathbb{E}X^2 = -\frac{q_1^2}{2\tilde{a}}.$$

To compute $\lim_{\epsilon\to 0}\mathbb{E}(X_{2,1}^2)$ first we set $\tilde{y} = a_{22}^{-1}a_{21}x + y$. Then, $(x, \tilde{y})$ is also an ergodic process with invariant distribution $\tilde{\rho}_\epsilon$ that converges as $\epsilon \to 0$ to $\mathcal{N}(0, \frac{q_1}{2\tilde{a}}) \otimes \mathcal{N}(0, \frac{q_2}{2a_{22}})$. Since $\xi(x,y) = -a_{21}x\tilde{y}$, it follows that

$$\lim_{\epsilon\to 0}\mathbb{E}_{\rho^\epsilon}(\xi(x,y)^2) = a_{12}^2 \frac{q_1}{2\tilde{a}}\frac{q_2}{2a_{22}}.$$

In addition, as $\epsilon \to 0$, the process $\tilde{y}$ decorrelates exponentially fast. Thus

$$\lim_{\epsilon\to 0}(P_t^\epsilon \xi)(x,y) = a_{12}\mathbb{E}(X(t)|X(0)=x)\mathbb{E}(\tilde{y}) \equiv 0$$

for all $t \geq 0$. As $t \to \infty$, the process $(x, \tilde{y})$ also converges exponentially fast to a mean-zero Gaussian distribution and thus the integral with respect to t is finite. We conclude that the second term of $\sigma(2,1)_\epsilon^2$ disappears as $\epsilon \to 0$ and thus

$$\lim_{\epsilon\to 0}\mathbb{E}(X_{2,1}^2) = a_{12}^2\frac{q_1 q_2}{4\tilde{a}a_{22}}.$$

Finally, we show that

$$\lim_{\epsilon\to 0}\mathbb{E}(X_{2,1}X_5) = 0.$$

Clearly, X_5 is independent of $\tilde{y}$ in the limit, since it only depends on x and U. So,

$$\lim_{\epsilon\to 0}\mathbb{E}(X_{2,1}X_5) = \lim_{\epsilon\to 0}\mathbb{E}\left(\mathbb{E}(X_{2,1}X_5|x)\right)$$

and

$$\lim_{\epsilon\to 0}\mathbb{E}\left(\mathbb{E}(X_{2,1}|x)\right) = 0$$

for the same reasons as above. Thus

$$\sigma_\epsilon^2 = \frac{4\tilde{a}^2}{q_1^2}\left(-\frac{q_1^2}{2\tilde{a}} + a_{12}^2\frac{q_1 q_2}{4\tilde{a}a_{22}}\right).$$

We have proved the following

Theorem 2.5. *Suppose that x is the projection to the x-coordinate of a solution of system* (1) *satisfying Assumptions 2.1. Let $\hat{a}_T^\epsilon$ be as in* (16)*. Then,*

$$\sqrt{T}\,(\hat{a}_T^\epsilon - \tilde{a}) \to \mathcal{N}(\mu_\epsilon, \sigma_\epsilon^2),$$

where

$$\mu_\epsilon \to 0 \quad \text{and} \quad \sigma_\epsilon^2 \to -2\tilde{a} + a_{12}^2\frac{\tilde{a}q_2}{a_{22}q_1} \quad \text{as} \quad \epsilon \to 0.$$

Remark 2.6. *Note that in the case where the data comes from the multiscale limit and for $\epsilon \to 0$, the asymptotic variance of the drift MLE (blue lines in Figure 1) is larger than that the asymptotic variance of the drift estimator where there is no misfit between model and data (red lines in Figure 1).*

2.4 The Diffusion Estimator

Suppose that we want to estimate the diffusion parameter of the process X described by (6) but we only observe a solution $\{x(t)\}_{t\in(0,T)}$ of (1a). As before, we replace X in the formula of the MLE (5) by x. In the following theorem, we show that the estimator is still consistent in the limit.

Theorem 2.7. *Suppose that x is the projection to the x-coordinate of a solution of system* (1) *satisfying Assumptions 2.1. We set*

$$\hat{q}_\delta^\epsilon = \frac{1}{T}\sum_{n=0}^{N-1}(x_{n+1}-x_n)^2 \tag{17}$$

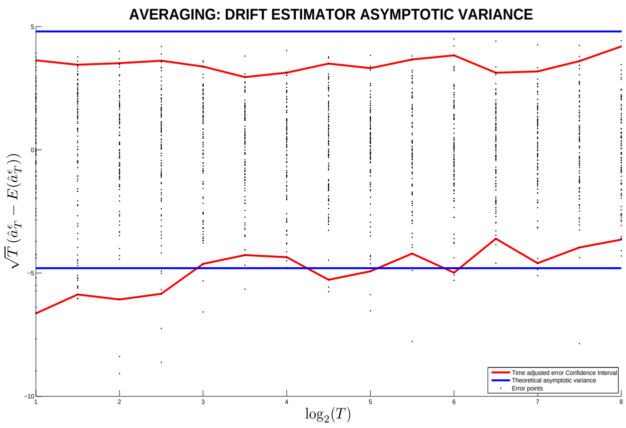

Figure 1: Averaging: Asymptotic Normality of $\hat{a}^\epsilon_T$

where $x_n = x(n\delta)$ is the discretized x process, $\delta \le \epsilon$ is the discretization step and $T = N\delta$ is fixed. Then, for every $\epsilon > 0$

$$\lim_{\delta\to 0} \mathbb{E}(\hat{q}^\epsilon_\delta - q_1)^2 = 0 \ ,$$

more specifically,

$$\mathbb{E}(\hat{q}^\epsilon_\delta - q_1)^2 = \mathcal{O}(\delta) \ .$$

Proof. We rewrite $x_{n+1} - x_n$ using discretized (1a),

$$x_{n+1} - x_n = \int_{n\delta}^{(n+1)\delta} \sqrt{q_1} dU_s + \hat{R}_1^{(n)} + \hat{R}_2^{(n)} \tag{18}$$

where

$$\begin{aligned} \hat{R}_1^{(n)} &= a_{11} \int_{n\delta}^{(n+1)\delta} x(s)ds \\ \hat{R}_2^{(n)} &= a_{12} \int_{n\delta}^{(n+1)\delta} y(s)ds \end{aligned}$$

We let $\xi_n = \frac{1}{\sqrt{\delta}} \left(U_{(n+1)\delta} - U_{n\delta}\right)$. We write

$$\int_{n\delta}^{(n+1)\delta} \sqrt{q_1} dU_s = \sqrt{q_1 \delta}\xi_n.$$

We can write the estimator as

$$\hat{q}^\epsilon_\delta = q_1 \frac{1}{N}\sum_{n=0}^{N-1} \xi_n^2 + 2\frac{\sqrt{q_1}}{N\sqrt{\delta}} \sum_{n=0}^{N-1} \xi_n (\hat{R}_1^{(n)} + \hat{R}_2^{(n)}) + \frac{1}{N\delta}\sum_{n=0}^{N-1} (\hat{R}_1^{(n)} + \hat{R}_2^{(n)})^2 \tag{19}$$

Hence, we can expand the error as

$$\mathbb{E}\left(\hat{q}_\delta^\epsilon - q_1\right)^2 \leq C\mathbb{E}\left(\frac{1}{N}\sum_{n=0}^{N-1}\xi_n^2 - 1\right)^2 \tag{20a}$$

$$+ C\frac{q_1}{N^2\delta}\mathbb{E}\left(\sum_{n=0}^{N-1}\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right)^2 \tag{20b}$$

$$+ C\frac{1}{N^2\delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})^2\right)^2 \tag{20c}$$

It is straightforward for line (20a),

$$\mathbb{E}\left(\frac{1}{N}\sum_{n=0}^{N-1}\xi_n^2 - 1\right)^2 = c\delta \ .$$

By Assumptions 2.1(v), and Hölder inequality, we have,

$$\begin{aligned}\mathbb{E}(\hat{R}_1^{(n)})^2 &= a_{11}^2\mathbb{E}\left(\int_{n\delta}^{(n+1)\delta} x(s)ds\right)^2 \\ &\leq ca_{11}^2\delta\int_{n\delta}^{(n+1)\delta}\mathbb{E}x(s)^2 ds \\ &\leq c\delta^2 \ .\end{aligned} \tag{21}$$

It is similar for $\mathbb{E}(\hat{R}_2^{(n)})^2$,

$$\begin{aligned}\mathbb{E}(\hat{R}_2^{(n)})^2 &= a_{12}^2\mathbb{E}\left(\int_{n\delta}^{(n+1)\delta} y(s)ds\right)^2 \\ &\leq ca_{12}^2\delta\int_{n\delta}^{(n+1)\delta}\mathbb{E}y(s)^2 ds \\ &\leq c\delta^2 \ .\end{aligned} \tag{22}$$

Since $\hat{R}_1^{(n)}$ and $\hat{R}_2^{(n)}$ are Gaussian random variables, we have $\mathbb{E}(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})^4 = C\delta^4$, so line (20c) is of order $\mathcal{O}(\delta^2)$. For line (20b), we need to get the correlation between $\hat{R}_i^{(n)}$ for $i \in \{1, 2\}$ and ξ_n. We write system (2) in integrated form,

$$x(s) = x_n + a_{11}\int_{n\delta}^{s} x(u)du + a_{12}\int_{n\delta}^{s} y(u)du + \sqrt{q_1}\int_{n\delta}^{s} dU_u \tag{23}$$

$$y(s) = y_n + \frac{a_{21}}{\epsilon}\int_{n\delta}^{s} x(u)du + \frac{a_{22}}{\epsilon}\int_{n\delta}^{s} y(u)du + \frac{\sqrt{q_2}}{\epsilon}\int_{n\delta}^{s} dV_u \tag{24}$$

We substitute (23) and (24) into $\hat{R}_1^{(n)}$ and $\hat{R}_2^{(n)}$ respectively,

$$\begin{aligned}
\hat{R}_1^{(n)} + \hat{R}_2^{(n)} &= \int_{n\delta}^{(n+1)\delta} a_{11}x(s) + a_{12}y(s)ds \\
&= a_{11}x_n\delta + a_{12}y_n\delta \\
&+ \left(a_{11}^2 + \frac{1}{\epsilon}a_{12}a_{21}\right)\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} x(u)duds \\
&+ \left(a_{11}a_{12} + \frac{1}{\epsilon}a_{12}a_{22}\right)\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} y(u)duds \\
&+ a_{11}\sqrt{q_1}\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} dU_u ds \\
&+ a_{12}\frac{\sqrt{q_2}}{\epsilon}\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} dV_u ds
\end{aligned}$$

Using this expansion, we find,

$$\begin{aligned}
& \mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right) \\
= \; & \mathbb{E}\left(\xi_n(a_{11}x_n\delta + a_{12}y_n\delta)\right) && \text{(25a)} \\
+ \; & \mathbb{E}\left(\xi_n\left(\left(a_{11}^2 + \frac{1}{\epsilon}a_{12}a_{21}\right)\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} x(u)duds\right)\right) && \text{(25b)} \\
+ \; & \mathbb{E}\left(\xi_n\left(a_{11}a_{12} + \frac{1}{\epsilon}a_{12}a_{22}\right)\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} y(u)duds\right) && \text{(25c)} \\
+ \; & \mathbb{E}\left(\xi_n\left(a_{11}\sqrt{q_1}\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} dU_u ds\right)\right) && \text{(25d)} \\
+ \; & \mathbb{E}\left(\xi_n\left(a_{12}\frac{\sqrt{q_2}}{\epsilon}\int_{n\delta}^{(n+1)\delta}\int_{n\delta}^{s} dV_u ds\right)\right) && \text{(25e)}
\end{aligned}$$

By the definition of ξ_n, line (25a) is zero. By substituting (23) and (24) into lines (25b) and (25c) respectively and iteratively, we know they are of orders $\mathcal{O}(\delta^2)$. By definition of ξ_n, we know that line (25d) is of order $\mathcal{O}(\delta^{\frac{3}{2}})$. By independence between U and V, line (25e) is zero. Therefore,

$$\mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right) = \mathcal{O}(\delta^{\frac{3}{2}}) \, .$$

Thus,

$$\mathbb{E}\left(\xi_n^2(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})^2\right) = \mathcal{O}(\delta^3) \, .$$

When $m < n$, we have,

$$\begin{aligned}
& \mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\xi_m(\hat{R}_1^{(m)} + \hat{R}_2^{(m)})\right) \\
= \; & \mathbb{E}\left(\mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\xi_m(\hat{R}_1^{(m)} + \hat{R}_2^{(m)})|\mathcal{F}_{n\delta}\right)\right) \\
= \; & \mathbb{E}\left(\xi_m(\hat{R}_1^{(m)} + \hat{R}_2^{(m)})\mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})|\mathcal{F}_{n\delta}\right)\right) \\
= \; & \mathbb{E}\left(\xi_m(\hat{R}_1^{(m)} + \hat{R}_2^{(m)})\right)\mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right) \\
= \; & \mathcal{O}(\delta^3) \, .
\end{aligned}$$

When $m > n$, the same result holds. Thus we have that line (20b) is of order $\mathcal{O}(\delta^2)$. Therefore, we have for equation (20),

$$\mathbb{E}\left(\hat{q}_\delta^\epsilon - q_1\right)^2 = \mathcal{O}(\delta) \, .$$

This completes the proof. □

2.5 Asymptotic Normality for the Diffusion Estimator

To examine the asymptotic normality of the diffusion estimator, we use the decomposition of $\hat{q}_\delta^\epsilon$ in the proof of Theorem 2.7,

$$\delta^{-\frac{1}{2}}(\hat{q}_\delta^\epsilon - q_1) = \delta^{-\frac{1}{2}} q_1 (\frac{1}{N}\sum_{n=0}^{N-1}\xi_n^2 - I) \tag{26a}$$

$$+ \; 2\delta^{-\frac{1}{2}} \frac{\sqrt{q_1}}{N\sqrt{\delta}} \sum_{n=0}^{N-1} \xi_n (\hat{R}_1^{(n)} + \hat{R}_2^{(n)}) \tag{26b}$$

$$+ \; \delta^{-\frac{1}{2}} \frac{1}{N\delta} \sum_{n=0}^{N-1} (\hat{R}_1^{(n)} + \hat{R}_2^{(n)})^2 \tag{26c}$$

Since

$$\lim_{\delta\to 0} \delta^{-\frac{1}{2}} q_1 (\frac{1}{N}\sum_{n=0}^{N-1}\xi_n^2 - I) = \lim_{N\to\infty} \frac{q_1}{\sqrt{T}} \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} (\xi_n^2 - I)$$

It follows from Central Limit Theorem for sum of multivariate i.i.d random variables, as $\delta \to 0$,

$$\lim_{\delta\to 0} \delta^{-\frac{1}{2}} q_1 (\frac{1}{N}\sum_{n=0}^{N-1}\xi_n^2 - I) \xrightarrow{D} \mathcal{N}(0, 2\frac{q_1^2}{T})$$

We have shown that $\mathbb{E}\left(\xi_n(\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right) = \mathcal{O}(\delta^{\frac{3}{2}})$, so line (26b) has mean

$$\mathbb{E}\left(\delta^{-\frac{1}{2}} \frac{\sqrt{q_1}}{N\sqrt{\delta}} \sum_{n=0}^{N-1} \xi_n (\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right) = \mathcal{O}(\delta^{\frac{1}{2}}) \, .$$

Using $\mathbb{E}\left(\sum_{n=0}^{N-1} \xi_n (\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right)^2 = \mathcal{O}(\delta)$, we find the second moment of (26b),

$$\mathbb{E}\left(\delta^{-\frac{1}{2}} \frac{\sqrt{q_1}}{N\sqrt{\delta}} \sum_{n=0}^{N-1} \xi_n (\hat{R}_1^{(n)} + \hat{R}_2^{(n)})\right)^2 = \mathcal{O}(\delta) \, .$$

Thus when δ is small,

$$\delta^{-\frac{1}{2}} \frac{\sqrt{q_1}}{N\sqrt{\delta}} \sum_{n=0}^{N-1} \xi_n (\hat{R}_1^{(n)} + \hat{R}_2^{(n)}) \sim \mathcal{N}(\mathcal{O}(\delta^{\frac{1}{2}}), \mathcal{O}(\delta)) \, .$$

Finally, for line (26c), using (21) and (22), we have

$$\mathbb{E}\left(\delta^{-\frac{1}{2}} \frac{1}{N\delta} \sum_{n=0}^{N-1} (\hat{R}_1^{(n)} + \hat{R}_2^{(n)})^2\right) = \mathcal{O}(\delta^{\frac{1}{2}}) \, ,$$

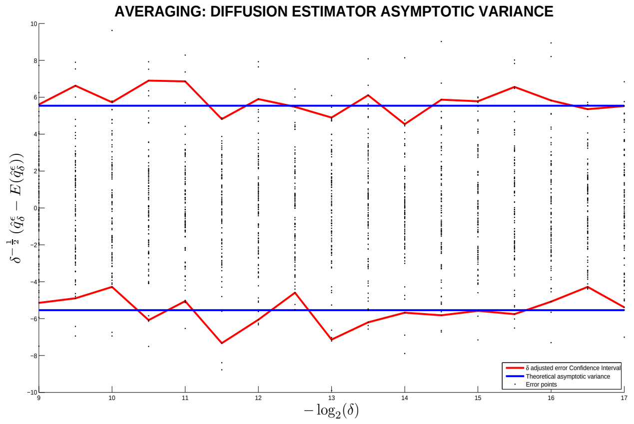

Figure 2: Averaging: Asymptotic Normality of $\hat{q}_\delta^\epsilon$

and,

$$\mathbb{E}\left(\delta^{-\frac{1}{2}}\frac{1}{N\delta}\sum_{n=0}^{N-1}(\hat{R}_1^{(n)}+\hat{R}_2^{(n)})^2\right)^2 = \mathcal{O}(\delta)\,.$$

Thus,

$$\delta^{-\frac{1}{2}}\frac{1}{N\delta}\sum_{n=0}^{N-1}(\hat{R}_1^{(n)}+\hat{R}_2^{(n)})^2 \sim \mathcal{N}(\mathcal{O}(\delta^{\frac{1}{2}}),\mathcal{O}(\delta))\,.$$

Putting all terms together, we have

$$\delta^{-\frac{1}{2}}\,(\hat{q}_\delta^\epsilon - q_1) \xrightarrow{D} \mathcal{N}(0,\frac{2q_1^2}{T})\,. \tag{27}$$

We have proved the following,

Theorem 2.8. *Under the conditions of Theorem 2.7 and with the same notation, it holds that*

$$\delta^{-\frac{1}{2}}\,(\hat{q}_\delta^\epsilon - q_1) \xrightarrow{D} \mathcal{N}(0,\frac{2q_1^2}{T}) \text{ as } \delta\to 0\,.$$

In Figure 2, we show an example of the distributions of the errors of the diffusion estimator as $\delta \to 0$.

3 Homogenization

We now consider the fast/slow system of stochastic differential equations described by (2), for the variables $(x,y)\in\mathcal{X}\times\mathcal{Y}$. We may take $\mathcal{X}$ and $\mathcal{Y}$ as either in $\mathbb{R}$ or $\mathbb{T}$. Our interest remains in data generated by the projection onto the x coordinate of the system.

Assumptions 3.1.
We assume that

(i) *U, V are independent Brownian motions;*

(ii) *q_1, q_2 are positive;*

(iii) $0 < \epsilon \ll 1$*;*

(iv) *the system's drift matrix*

$$\begin{pmatrix} \frac{1}{\epsilon}a_{11} + a_{13} & \frac{1}{\epsilon}a_{12} + a_{14} \\ \frac{1}{\epsilon^2}a_{21} & \frac{1}{\epsilon^2}a_{22} \end{pmatrix}$$

only have negative real eigenvalues when ϵ is sufficiently small;

(v) $a_{21} \neq 0$*;*

(vi) *$x(0)$ and $y(0)$ are independent of U and V, $(x(0), y(0))$ is under the invariant measure of system* (1)*, and* $\mathbb{E}\left(x^2(0) + y^2(0)\right) < \infty$.

Remark 3.2. *In assumption 3.1(iv), we have assumed the whole system* (2) *to be ergodic when ϵ is sufficiently small. This condition can be decomposed to a_{22} and $a_{13} - a_{14}a_{22}^{-1}a_{21}$ are negative real numbers; and $a_{11} - a_{12}a_{22}^{-1}a_{21} = 0$, which ensures the fast scale term in (2a) vanishes.*

Remark 3.3. *Assumption 3.1(v) is necessary in our setup, however, the result could still hold when a_{21} is zero, an example is discussed by Papavasiliou in [10] for diffusion estimates.*

Under assumptions 3.1, the solution (x, y) of (2) is ergodic. In addition, x converges as $\epsilon \to 0$ to the solution of the homogenized equation

$$\frac{dX}{dt} = \tilde{a}X + \sqrt{\tilde{q}}\frac{dW}{dt} \tag{28}$$

where

$$\tilde{a} = a_{13} - a_{14}a_{22}^{-1}a_{21} \tag{29}$$

and

$$\tilde{q} = q_1 + a_{12}^2 a_{22}^{-2} q_2 \tag{30}$$

The convergence of the homogenizing systems is different from that of the averaging systems. For each given time series of observations, the paths of the slow process converge to the paths of the corresponding homogenized equation. However, we will see that in the limit $\epsilon \to 0$, the likelihood of the drift or diffusion parameter is different depending on whether we observe a path of the slow process generated by (2a) or the homogenized process (28) (see also [23, 25, 26]).

3.1 The Paths

The following theorem extends Theorem 18.1 in [25], which gives weak convergence of paths on $\mathbb{T}$. By limiting ourselves to the OU process, we extend the domain to $\mathbb{R}$ and prove a stronger mode of convergence.

Lemma 3.4. *Suppose that (x, y) solves (2a) and Assumptions 3.1 are satisfied. Then, for fixed finite $T > 0$ and small ϵ,*

$$\mathbb{E} \sup_{0 \le t \le T} \left(x^2(t) + y^2(t)\right) = \mathcal{O}\left(\log(1 + \frac{T}{\epsilon^2})\right). \tag{31}$$

Proof. We look at the system of SDEs as,

$$d\mathbf{x}_t = \mathbf{a}\mathbf{x}_t dt + \sqrt{\mathbf{q}} dW_t \tag{32}$$

where,

$$\mathbf{x} = \begin{pmatrix} x \\ y \end{pmatrix}, \mathbf{a} = \begin{pmatrix} \frac{1}{\epsilon}a_{11} + a_{13} & \frac{1}{\epsilon}a_{12} + a_{14} \\ \frac{1}{\epsilon^2}a_{21} & \frac{1}{\epsilon^2}a_{22} \end{pmatrix} \text{ and } \mathbf{q} = \begin{pmatrix} q_1 & 0 \\ 0 & \frac{1}{\epsilon^2}q_2 \end{pmatrix}.$$

We try to characterize the magnitude of the eigenvalues of $\mathbf{a}$. To find the eigenvalues, we require

$$\det(\mathbf{a} - \lambda I) = 0 \,.$$

By solving this system and using existing results regarding the eigenvalues of a perturbed matrix of $\tilde{a}$ (see [14][p. 137, Theorem 2]), we find that the eigenvalues will be of order $\mathcal{O}(1)$ and $\mathcal{O}(1/\epsilon^2)$. Therefore, we can decompose $\mathbf{a}$ as

$$\mathbf{a} = PDP^{-1} \text{ with } D = \begin{pmatrix} D_1 & 0 \\ 0 & \frac{1}{\epsilon^2}D_2 \end{pmatrix}$$

where D is the diagonal matrix, for which $D_1 \in \mathbb{R}$ and $D_2 \in \mathbb{R}$ are diagonal entries of order $\mathcal{O}(1)$. Following exactly the same approach as in lemma 2.2, we get the result. □

Theorem 3.5. *Let Assumptions 3.1 hold for system* (2)*. Suppose that x and X are solutions of (2a) and* (28) *respectively. (x, y) corresponds to the realization (U, V) of Brownian motion, while X corresponds to the realization*

$$W_\cdot = \tilde{q}^{-\frac{1}{2}} \left(\sqrt{q_1} U - a_{12} a_{22}^{-1} \sqrt{q_2} V \right) \tag{33}$$

and $x(0) = X(0)$. Then x converges to X in L^2. More specifically,

$$\mathbb{E} \sup_{0 \le t \le T} (x(t) - X(t))^2 \le c \left(\epsilon^2 \log(\frac{T}{\epsilon}) + \epsilon^2 T \right) e^T \,,$$

when T is fixed finite, the above bound can be simplified to

$$\mathbb{E} \sup_{0 \le t \le T} (x(t) - X(t))^2 = \mathcal{O}(\epsilon^2 \log(\epsilon)).$$

Proof. We rewrite (2b) as

$$(a_{22}^{-1} a_{21} x(t) + y(t)) dt = \epsilon^2 a_{22}^{-1} dy(t) - \epsilon a_{22}^{-1} \sqrt{q_2} dV_t \,. \tag{34}$$

We also rewrite (2a) as

$$\begin{aligned} dx(t) &= \frac{1}{\epsilon} a_{12} (a_{22}^{-1} a_{21} x(t) + y(t)) dt + a_{14} (a_{22}^{-1} a_{21} x(t) + y(t)) dt \\ &\quad + (a_{13} - a_{14} a_{22}^{-1} a_{21}) x(t) dt + \sqrt{q_1} dU_t \\ &= \left(\frac{1}{\epsilon} a_{12} + a_{14} \right) (a_{22}^{-1} a_{21} x(t) + y(t)) dt \\ &\quad + \tilde{a} x(t) dt + \sqrt{q_1} dU_t \,. \end{aligned} \tag{35}$$

Replacing $(a_{22}^{-1}a_{21}x(t)+y(t))dt$ in (35) by the right-hand-side of (34), we get

$$\begin{aligned} dx(t) &= \epsilon(a_{12}+\epsilon a_{14})a_{22}^{-1}dy(t) - a_{12}a_{22}^{-1}\sqrt{q_2}dV_t - \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}dV_t \\ &\quad + \tilde{a}x(t)dt + \sqrt{q_1}dU_t \\ &= \tilde{a}x(t)dt + \epsilon(a_{12}+\epsilon a_{14})a_{22}^{-1}dy(t) \\ &\quad + \sqrt{\tilde{q}}dW_t - \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}dV_t \, . \end{aligned} \tag{36}$$

Thus

$$\begin{aligned} x(t) &= x(0) + \int_0^t \tilde{a}x(s)ds + \sqrt{\tilde{q}}W_t \\ &\quad + \epsilon(a_{12}+\epsilon a_{14})a_{22}^{-1}(y(t)-y(0)) - \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}V_t \, . \end{aligned} \tag{37}$$

Recall that the homogenized equation (28) is

$$X(t) = X(0) + \int_0^t \tilde{a}X(s)ds + \sqrt{\tilde{q}}W_t \, . \tag{38}$$

Let $e(t) = x(t) - X(t)$. Subtracting the previous equation from (37) and using the assumption $X(0) = x(0)$, we find that

$$\begin{aligned} e(t) &= \tilde{a}\int_0^t e(s)ds \\ &\quad + \epsilon\left((a_{12}+\epsilon a_{14})a_{22}^{-1}(y(t)-y(0)) - a_{14}a_{22}^{-1}\sqrt{q_2}V_t\right) \, . \end{aligned} \tag{39}$$

Applying Lemma 3.4, we find an ϵ-independent constant C, such that

$$\mathbb{E}\left(\sup_{0\le t\le T} y^2(t)\right) \le C\log(\frac{T}{\epsilon}) \, .$$

By Cauchy-Schwarz,

$$\mathbb{E}\left(\sup_{0\le t\le T} e^2(t)\right) \le c\left(\int_0^T \mathbb{E}e^2(s)ds + \epsilon^2\log(\frac{T}{\epsilon}) + \epsilon^2 T\right) \, . \tag{40}$$

By the integrated version of the Gronwall inequality [25], we deduce that

$$\mathbb{E}\left(\sup_{0\le t\le T} e^2(t)\right) \le c\left(\epsilon^2\log(\frac{T}{\epsilon}) + \epsilon^2 T\right)e^T \, . \tag{41}$$

When T is finite, we have

$$\mathbb{E}\left(\sup_{0\le t\le T} e^2(t)\right) = \mathcal{O}\left(\epsilon^2\log(\epsilon)\right) \, .$$

This completes the proof. □

3.2 The Drift Estimator

As in the averaging case, a natural idea for estimating the drift of the homogenized equation is to use the maximum likelihood estimator (4), replacing X by the solution x

of (2a). However, in the case of homogenization we do not get asymptotically consistent estimates. To achieve this, we must subsample the data: we choose Δ (time step for observations) according to the value of the scale parameter ϵ and solve the estimation problem for discretely observed diffusions (see [23, 25, 26]). The maximum likelihood estimator for the drift of a homogenized equation converges after proper subsampling. We let the observation time interval Δ and the number of observations N both depend on the scaling parameter ϵ, by setting $\Delta = \epsilon^\alpha$ and $N = \epsilon^{-\gamma}$. We find the error is optimized in the L^2 sense when $\alpha = 1/2$. We will show that $\hat{a}_{N,\epsilon}$ converges to $\tilde{a}$ only if $\frac{\Delta}{\epsilon^2} \to \infty$, in a sense to be made precise later.

Theorem 3.6. *Suppose that x is the projection to the x-coordinate of a solution of system* (2) *satisfying Assumptions 3.1. Let $\hat{a}_{N,\epsilon}$ be the estimate we get by replacing X in* (4) *by x, i.e.*

$$\hat{a}_{N,\epsilon} = \left(\frac{1}{N\Delta} \sum_{n=0}^{N-1} x_n \left(x_{n+1} - x_n\right) \right) \left(\frac{1}{N\Delta} \sum_{n=0}^{N-1} x_n^2 \Delta \right)^{-1} \tag{42}$$

Then,

$$\mathbb{E}(\hat{a}_{N,\epsilon} - \tilde{a})^2 = \mathcal{O}(\Delta + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2})$$

where $\tilde{a}$ as defined in (29)*. Consequently, if $\Delta = \epsilon^\alpha$, $N = \epsilon^{-\gamma}$, $\alpha \in (0, 1)$, $\gamma > \alpha$,*

$$\lim_{\epsilon \to 0} \mathbb{E}(\hat{a}_{N,\epsilon} - \tilde{a})^2 = 0 \ .$$

Furthermore, $\alpha = 1/2$ and $\gamma \geq 3/2$ optimize the error.

Before proving Theorem 3.6, we first find the magnitude of the increment of y over a small time interval Δ. Solving equation (2b), we have

$$\begin{aligned} y_{n+1} - y_n &= (e^{a_{22}\frac{\Delta}{\epsilon^2}} - I)y_n \\ &+ \frac{1}{\epsilon^2} \int_{n\Delta}^{(n+1)\Delta} e^{a_{22}\frac{(n+1)\Delta - s}{\epsilon^2}} x(s)ds \\ &+ \frac{1}{\epsilon} \int_{n\Delta}^{(n+1)\Delta} e^{a_{22}\frac{(n+1)\Delta - s}{\epsilon^2}} \sqrt{q_2} dV_s \ . \end{aligned} \tag{43}$$

By triangle inequality, we have

$$\begin{aligned} \mathbb{E}(y_{n+1} - y_n)^2 &\leq (e^{a_{22}\frac{\Delta}{\epsilon^2}} - 1)\mathbb{E}y_n^2 \\ &+ c(e^{a_{22}\frac{\Delta}{\epsilon^2}} - 1) \\ &+ \frac{1}{2}(e^{2a_{22}\frac{\Delta}{\epsilon^2}})^2 q_2^2 \ . \end{aligned}$$

Since a_{22} is a negative constant,

$$\mathbb{E}(y_{n+1} - y_n)^2 = \mathcal{O}(e^{-\frac{\Delta}{\epsilon^2}} - 1) \ .$$

By definition $\Delta = \epsilon^\alpha$, and the property that $(e^{-\frac{\Delta}{\epsilon^2}} - 1) = \mathcal{O}(\frac{\Delta}{\epsilon^2})$ if $\frac{\Delta}{\epsilon^2}$ is small, the above equation can be rewritten as

$$\mathbb{E}(y_{n+1} - y_n)^2 = \mathcal{O}(\epsilon^{\max(\alpha - 2, 0)}) \ . \tag{44}$$

Proof. Define I_1 and I_2 as

$$I_1 = \frac{1}{N\Delta}\sum_{n=0}^{N-1}(x_{n+1}-x_n)x_n \ , \ I_2 = \frac{1}{N\Delta}\sum_{n=0}^{N-1}x_n^2\Delta$$

By ergodic theorem, and since $N = \epsilon^{-\gamma}$, we have

$$\lim_{\epsilon\to 0} I_2 = \mathbb{E}\left(x_n^2\right) = C \neq 0$$

which is a non-zero constant. Hence instead of proving

$$\mathbb{E}(\hat{a}_{N,\epsilon}-\tilde{a})^2 = \mathcal{O}(\Delta^2 + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2}) \ ,$$

we prove,

$$\mathbb{E}(I_1 - \tilde{a}I_2)^2 = \mathcal{O}(\Delta^2 + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2}) \ .$$

We use the rearranged equation (36) of (2a) to decompose the error,

$$I_1 - \tilde{a}I_2 = J_1 + J_2 + J_3 + J_4 \ . \tag{45}$$

where

$$\begin{aligned}
J_1 &= \frac{1}{N\Delta}\sum_{n=0}^{N-1}\left(\tilde{a}\int_{n\Delta}^{(n+1)\Delta} x(s)ds - x_n\right)x_n \\
J_2 &= \frac{1}{N\Delta}\sum_{n=0}^{N-1}\left(\sqrt{\tilde{q}}\int_{n\Delta}^{(n+1)\Delta} x_n dW_s\right) \\
J_3 &= \frac{\epsilon}{N\Delta}\sum_{n=0}^{N-1}(a_{12}+\epsilon a_{14})a_{22}^{-1}\int_{n\Delta}^{(n+1)\Delta} x_n dy(s) \\
J_4 &= \frac{\epsilon}{N\Delta}\sum_{n=0}^{N-1}a_{14}a_{22}^{-1}\sqrt{q_2}\int_{n\Delta}^{(n+1)\Delta} x_n dV_s
\end{aligned}$$

By independence, Itô isometry and ergodicity, we immediately have

$$\begin{aligned}
\mathbb{E}J_2^2 &= \mathbb{E}\left(\frac{\sqrt{\tilde{q}}}{N\Delta}\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x_n dW_s\right)^2 \\
&= \frac{\tilde{q}}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x_n dW_s\right)^2 \\
&\leq \frac{\tilde{q}}{N^2\Delta^2}N\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta} dW_s\right)^2 \mathbb{E}x_n^2 \\
&\leq \frac{\tilde{q}}{N^2\Delta^2}N\Delta\mathbb{E}x_n^2 \\
&= \mathcal{O}(\frac{1}{N\Delta}) \ ,
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}J_4^2 &= \mathbb{E}\left(\frac{\epsilon C}{N\Delta}\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x_n dV_s\right)^2 \\
&= \frac{\epsilon^2 C}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x_n dV_s\right)^2 \\
&\leq \frac{\epsilon^2 C}{N^2\Delta^2} N\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta} dV_s\right)^2 \mathbb{E}(x_n^2) \\
&\leq \frac{\epsilon^2 C}{N^2\Delta^2} N\Delta\mathbb{E}(x_n^2) \\
&= \mathcal{O}(\frac{\epsilon^2}{N\Delta})\,.
\end{aligned}$$

By Hölder inequality, and (44), we have,

$$\begin{aligned}
\mathbb{E}J_3^2 &= \mathbb{E}\left(\frac{\epsilon C}{N\Delta}\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta} x_n dy\right)^2 \\
&= \mathbb{E}\left(\frac{\epsilon C}{N\Delta}\sum_{n=0}^{N-1} x_n(y_{n+1}-y_n)\right)^2 \\
&\leq \frac{\epsilon^2}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}(y_{n+1}-y_n)\right)^2 \mathbb{E}(\sum_{n=0}^{N-1} x_n)^2 \\
&\leq \frac{\epsilon^2 C}{N^2\Delta^2} N(\epsilon^{\max(\alpha-2,0)})N\mathbb{E}x_n^2 \\
&= \mathcal{O}(\frac{\epsilon^2}{\Delta^2})\,.
\end{aligned}$$

Finally, we find the squared error for J_1. We use the integrated form of equation (36) on time interval $[n\Delta, s]$ to replace $x(s)$

$$\mathbb{E}J_1^2 = \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}(x(s)-x_n)x_n ds\right)^2 \tag{46}$$

$$= \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\left(\sum_{n=0}^{N-1}(K_1^{(n)}+K_2^{(n)}+K_3^{(n)}+K_4^{(n)})\right)^2 \tag{47}$$

(48)

where,

$$\begin{aligned}
K_1^{(n)} &= \tilde{a}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_n x(u)duds\ ,\\
K_2^{(n)} &= \epsilon(a_{12}+\epsilon a_{14})a_{22}^{-1}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_n dy(u)ds\ ,\\
K_3^{(n)} &= \sqrt{\tilde{q}}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_n dW_u ds\ ,\\
K_4^{(n)} &= \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_n dV_u ds\ .
\end{aligned}$$

We immediately see that

$$\mathbb{E}J_1^2 = \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\sum_{n=0}^{N-1}\left(\sum_{i=1}^{4}K_i^{(n)}\right)^2 \tag{49}$$

$$+ \frac{\tilde{a}^2}{N^2\Delta^2}\mathbb{E}\sum_{m\neq n}|\left(\sum_{i=1}^{4}K_i^{(n)}\right)\left(\sum_{j=1}^{4}K_j^{(m)}\right)| \tag{50}$$

Remark 3.7. *Under the vector valued problem, we use the exact decomposition of* $\mathbb{E}\|J_1\|^2$ *by using* (49) *and* (50)*. This is essential in order to obtain more optimized subsampling rate for the drift estimator. For general* L^p *bound for the error, we can apply Hölder's inequality to decompose* J_1 *as,*

$$\begin{aligned}
\mathbb{E}\|J_1\|^p &= \frac{C}{N^p\Delta^p}\mathbb{E}\|\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}(x(s)-x_n)ds\otimes x_n\|^p\\
&\leq \frac{C}{N^{p-1}\Delta^p}\mathbb{E}\|\sum_{n=0}^{N-1}\int_{n\Delta}^{(n+1)\Delta}(x(s)-x_n)ds\|^p\mathbb{E}\|x_n\|^p
\end{aligned}$$

which is used in [26]. Using this inequality will give an optimal subsampling rate of $\alpha = 2/3$*, and achieves an over all* L^1 *error of order* $\mathcal{O}(\epsilon^{1/3})$*. However, this magnitude of overall error is not optimal in* L^2*. We will show later that the optimal* L^2 *error can be achieved at the order of* $\mathcal{O}(\epsilon^{1/2})$*, using the exact decomposition shown above.*

By Cauchy-Schwarz inequality, we know for line (49),

$$\mathbb{E}\sum_{n=0}^{N-1}\left(\sum_{i=1}^{4}K_i^{(n)}\right)^2 \leq \sum_{n=0}^{N-1}\sum_{i=1}^{4}\mathbb{E}\left(K_i^{(n)}\right)^2\ .$$

Using first order iterated integrals, we have

$$\begin{aligned}
\mathbb{E}(K_1^{(n)})^2 &= \mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_n x(u)duds\right)^2\\
&\leq C\Delta\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x^2(u)dudsx_n^2\\
&\leq C\Delta\int_{n\Delta}^{(n+1)\Delta}(s-n\Delta)^2ds\\
&= \mathcal{O}(\Delta^4)\ .
\end{aligned}$$

Using (44), we have

$$\begin{aligned}
\mathbb{E}(K_2^{(n)}ds)^2 &= \mathbb{E}\left(\epsilon C\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x_n dy(u)ds\right) \\
&\le C\epsilon^2\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta} x_n(y(s)-y(u))ds\right)^2 \\
&\le C\epsilon^2\Delta\mathbb{E}\int_{n\Delta}^{(n+1)\Delta}(y(s)-y(u))^2 ds x_n^2 \\
&\le C\epsilon^2\Delta\mathbb{E}\int_{n\Delta}^{(n+1)\Delta}(e^{-\frac{s-n\Delta}{\epsilon^2}}-1)ds \\
&= \mathcal{O}\left(\epsilon^4(e^{-\frac{\Delta}{\epsilon^2}}-1)\right) .
\end{aligned}$$

For $K_3^{(n)}$, we have,

$$\begin{aligned}
\mathbb{E}(K_3^{(n)})^2 &= \mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s}\sqrt{\tilde{q}}x_n dW_u ds\right)^2 \\
&\le C\Delta\int_{n\Delta}^{(n+1)\Delta}\left(\int_{n\Delta}^{s} dW_u\right)^2 ds \\
&\le C\Delta\int_{n\Delta}^{(n+1)\Delta}(s-n\Delta)ds \\
&= \mathcal{O}(\Delta^3) .
\end{aligned}$$

Since $K_4^{(n)}$ is similar to $K_3^{(n)}$, we have

$$\mathbb{E}(K_4^{(n)})^2 = \mathcal{O}(\epsilon^2\Delta^3) .$$

Thus, for line (49), the order of the dominating terms are,

$$\mathbb{E}\sum_{n=0}^{N-1}\left(\sum_{i=1}^{4}K_i^{(n)}\right)^2 = \mathcal{O}(N\Delta^4 + N\epsilon^4(e^{-\frac{\Delta}{\epsilon^2}}-1) + N\Delta^3) .$$

For line (50),

$$\mathbb{E}\sum_{m\neq n}|(\sum_{i=1}^{4}K_i^{(n)})(\sum_{j=1}^{4}K_j^{(m)})| \le \sum_{m\neq n}\mathbb{E}|\sum_{i=1}^{4}K_i^{(n)}|\mathbb{E}|\sum_{j=1}^{4}K_j^{(m)}| .$$

We know,

$$\begin{aligned}
\mathbb{E}(K_1^{(n)}) &= \mathbb{E}|C\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} x(u)duds| \\
&\le C\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}(s-n\Delta)ds\right) \\
&= \mathcal{O}(\Delta^2) .
\end{aligned}$$

Similarly, we have

$$\begin{aligned}\mathbb{E}|K_2^{(n)}| &= \epsilon C\mathbb{E}\left(\int_{n\Delta}^{(n+1)\Delta}(y(s)-y_n)ds\right)\\ &= \mathcal{O}(\epsilon\Delta)\,.\end{aligned}$$

Since the integral of Brownian motions is Gaussian

$$\begin{aligned}\mathbb{E}|K_3^{(n)}| &= C\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} dW_u ds)\\ &= C\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}(W(s)-W(n\Delta))ds)\\ &= C\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}W(s)ds - W(n\Delta)\Delta)\\ &= 0\,.\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}|K_4^{(n)}| &= C\epsilon\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}\int_{n\Delta}^{s} dV_u ds)\\ &= C\epsilon\mathbb{E}(\int_{n\Delta}^{(n+1)\Delta}V(s)ds - V(n\Delta)\Delta)\\ &= 0\,.\end{aligned}$$

Thus,

$$\mathbb{E}|\sum_{i=1}^{4}K_i^{(n)}| = \mathcal{O}(\Delta^2+\epsilon\Delta)\,,$$

immediately we have for line (50),

$$\mathbb{E}\sum_{m\neq n}|(\sum_{i=1}^{4}K_i^{(n)})(\sum_{j=1}^{4}K_j^{(m)})| = \mathcal{O}(N^2\Delta^4+N^2\epsilon^2\Delta^2)\,.$$

Putting all terms for J_1 together, we keep the dominating terms, and by assumption $N\Delta\to\infty$, and $\alpha<2$ since $e^{-\frac{\Delta}{\epsilon^2}}\to 0$,

$$\begin{aligned}\mathbb{E}J_1^2 &\leq \frac{C}{N^2\Delta^2}(N\Delta^4+N\epsilon^4(e^{-\frac{\Delta}{\epsilon^2}}-1)+N\Delta^3)\\ &+ \frac{C}{N^2\Delta^2}(N^2\Delta^4+N^2\epsilon^2\Delta^2)\\ &= \mathcal{O}(\frac{\Delta^2}{N}+\frac{\epsilon^4}{N\Delta^2}(e^{-\frac{\Delta}{\epsilon^2}}-1)+\frac{\Delta}{N}+\Delta^2+\epsilon^2)\\ &= \mathcal{O}(\frac{\epsilon^4}{N\Delta^2}+\Delta^2+\epsilon^2)\,.\end{aligned}$$

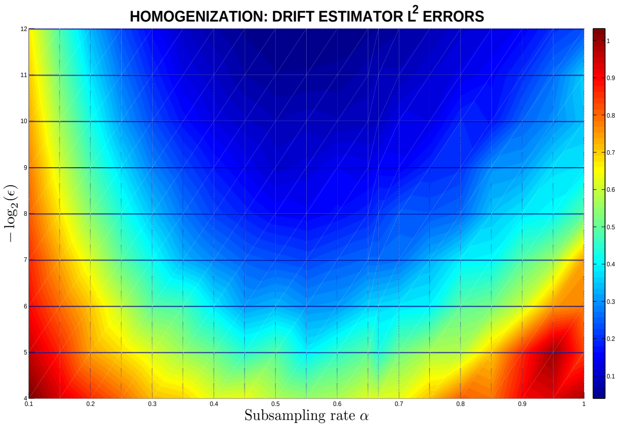

Figure 3: Homogenization: L^2 norm of $(\hat{a}_{N,\epsilon} - \tilde{a})$ for different ϵ and α

Therefore, putting J_i's, $i \in \{1, 2, 3, 4\}$, together, we have,

$$\begin{aligned} \mathbb{E}(I_1 - \tilde{a}I_2)^2 &\leq \sum_{i=1}^{4} \mathbb{E}J_i^2 \\ &= \mathcal{O}(\frac{\epsilon^4}{N\Delta^2} + \Delta^2 + \epsilon^2) \\ &+ \mathcal{O}(\frac{1}{N\Delta}) \\ &+ \mathcal{O}(\frac{\epsilon^2}{\Delta^2}) \\ &+ \mathcal{O}(\frac{\epsilon^2}{N\Delta}) \\ &= \mathcal{O}(\Delta^2 + \frac{1}{N\Delta} + \frac{\epsilon^2}{\Delta^2}) \end{aligned}$$

We rewrite the above equation using $\Delta = \epsilon^\alpha$ and $N = \epsilon^{-\gamma}$,

$$\mathbb{E}(I_1 - \tilde{a}I_2)^2 = \mathcal{O}(\epsilon^{2\alpha} + \epsilon^{\gamma-\alpha} + \epsilon^{2-2\alpha}) .$$

It is immediately seen that $\alpha = \frac{1}{2}$ and $\gamma \geq 3/2$ optimize the error, and $\alpha \in (0, 1)$, the order of the error is

$$\mathbb{E}(I_1 - \tilde{a}I_2)^2 = \mathcal{O}(\epsilon) .$$

This completes the proof. □

In Figure 3, we show an example of the L^2 error of the drift estimator with various scaling parameter ϵ and subsampling rate α. We see that the error is minimized around $\alpha = 1/2$ as in Theorem 3.6.

3.3 The Diffusion Estimator

Just as in the case of the drift estimator, we define the diffusion estimator by the maximum likelihood estimator (5), where X is replaced by the discretized solution of (2a).

More specifically, we define

$$\tilde{q}^{\epsilon}_{N,\Delta} = \frac{1}{N\Delta}\sum_{n=0}^{N-1}(x_{n+1}-x_n)^2 \tag{51}$$

where $x_n = x(n\Delta)$ is the discrete observation of the process generated by (2a) and Δ is the observation time interval.

Theorem 3.8. *Suppose that x is the projection to the x-coordinate of a solution of system* (2) *satisfying Assumptions 3.1. Let $\hat{q}_\epsilon$ be the estimate we get by replacing X in* (5) *by x, i.e.*

$$\hat{q}_\epsilon = \frac{1}{T}\sum_{n=0}^{N-1}(x_{n+1}-x_n)^2 \; .$$

Then

$$\mathbb{E}(\hat{q}_\epsilon - \tilde{q})^2 = \mathcal{O}\left(\Delta + \epsilon^2 + \frac{\epsilon^4}{\Delta^2}\right)$$

where $\tilde{q}$ as defined in (30)*. Consequently, if $\Delta = \epsilon^\alpha$, fix $T = N\Delta$, and $\alpha \in (0,2)$, then*

$$\lim_{\epsilon\to 0}\mathbb{E}(\hat{q}_\epsilon - \tilde{q})^2 = 0 \; .$$

Furthermore, $\alpha = 4/3$ optimizes the error.

We first define

$$\sqrt{\Delta}\eta_n = \int_{n\Delta}^{(n+1)\Delta} dW_t \; .$$

Proof. We now prove Theorem 3.8. Using the integral form of equation (36),

$$\begin{aligned} x_{n+1} - x_n &= \int_{n\Delta}^{(n+1)\Delta} \sqrt{\tilde{q}}dW_s \\ &+ \hat{R}_1 + \hat{R}_2 + \hat{R}_3 \end{aligned} \tag{52}$$

where

$$\begin{aligned} \hat{R}_1 &= \tilde{a}\int_{n\Delta}^{(n+1)\Delta} x(s)ds \\ \hat{R}_2 &= \epsilon a_{14}a_{22}^{-1}\sqrt{q_2}\int_{n\Delta}^{(n+1)\Delta} dV_s \\ \hat{R}_3 &= \epsilon(a_{12}+\epsilon a_{14})a_{22}^{-1}\int_{n\Delta}^{(n+1)\Delta} dy(s) \end{aligned}$$

We rewrite line (52) as

$$\int_{n\Delta}^{(n+1)\Delta} \sqrt{\tilde{q}}dW_s = \sqrt{\tilde{q}\Delta}\eta_n$$

where η_n are $\mathcal{N}(0,1)$ random variables.

For Δ and ϵ sufficiently small, by Cauchy-Schwarz inequality

$$\begin{aligned}
\mathbb{E}\left(c\int_{n\Delta}^{(n+1)\Delta} x(s)ds\right)^2 &\leq c\mathbb{E}\int_{n\Delta}^{(n+1)\Delta} x^2(s)ds\int_{n\Delta}^{(n+1)\Delta} ds \\
&\leq c\Delta\mathbb{E}\int_{n\Delta}^{(n+1)\Delta} x^2(s)ds \\
&\leq c\Delta^2\mathbb{E}\left(\sup_{n\Delta\leq s\leq(n+1)\Delta} x^2(s)\right) \\
&= \mathcal{O}(\Delta^2)
\end{aligned}$$

Therefore,

$$\mathbb{E}(\hat{R}_1)^2 = \mathcal{O}(\Delta^2)$$

By Itô isometry

$$\mathbb{E}(\hat{R}_2)^2 = \mathcal{O}(\epsilon^2\Delta)$$

Then we look at $\hat{R}_3$,

$$\mathbb{E}(\hat{R}_3)^2 = \epsilon^2 C\mathbb{E}(y_{n+1}-y_n)^2$$

By (44), we have

$$\mathbb{E}(\hat{R}_3)^2 = \mathcal{O}(\epsilon^{\max(\alpha,2)}) \tag{53}$$

We substitute $(x_{n+1}-x_n)$ into the estimator $\hat{q}_\epsilon$ in Theorem 3.8. We decompose the estimator's error as follows,

$$\begin{aligned}
\hat{q}_\epsilon - \tilde{q} &= \tilde{q}(\frac{1}{N}\sum_{n=0}^{N-1}\eta_n^2 - 1) \\
&+ \frac{1}{T}\sum_{n=0}^{N-1}\sum_{i=1}^{3}\left(\hat{R}_i^2\right) \\
&+ \frac{2}{T}\sum_{n=0}^{N-1}\sum_{i=1}^{3}\hat{R}_i\sqrt{\tilde{q}\Delta}\eta_n \\
&+ \frac{1}{T}\sum_{n=0}^{N-1}\left(\sum_{i\neq j}\hat{R}_i\hat{R}_j\right) \\
&= R
\end{aligned}$$

Then we bound the mean squared error using Cauchy-Schwarz inequality.

$$\begin{aligned}
\mathbb{E}\left(\hat{q}_\epsilon - \tilde{q}\right)^2 &\leq C\tilde{q}^2\mathbb{E}(\frac{1}{N}\sum_{n=0}^{N-1}\eta_n^2 - 1)^2 &(54)\\
&+ C\sum_{i=1}^{3}\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_i^2\right)^2 &(55)\\
&+ C\sum_{i=1}^{3}\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_i\sqrt{\tilde{q}\Delta}\eta_n\right)^2 &(56)\\
&+ C\sum_{i\neq j}\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\left(\hat{R}_i\otimes\hat{R}_j\right)\right)^2 &(57)
\end{aligned}$$

By law of large numbers, line (54) is of order $\mathcal{O}(\Delta)$.

In line (55), for $i \in \{1, 2\}$, we have

$$\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_i^2\right)^2 = \frac{1}{T^2}N\sum_{n=0}^{N-1}\mathbb{E}(\hat{R}_i^2)^2.$$

Since $\mathbb{E}(\hat{R}_1)^2 = \mathcal{O}(\Delta^2)$, we have

$$\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_1^2\right)^2 = \mathcal{O}\left(N^2(\Delta^2)^2\right) = \mathcal{O}\left(\Delta^2\right);$$

since $\mathbb{E}(\hat{R}_2)^2 = \mathcal{O}(\epsilon^2\Delta)$, we have

$$\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_2^2\right)^2 = \mathcal{O}\left(N^2(\Delta\epsilon^2)^2\right) = \mathcal{O}(\epsilon^4).$$

It is different for $\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_3^2\right)^2$, by (44), we have

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_3^2\right)^2 &= \frac{C\epsilon^4}{T^2}\mathbb{E}\left(\sum_{n=0}^{N-1}(y_{n+1}-y_n)^2\right)^2 \\
&\leq C\epsilon^4 N\sum_{n=0}^{N-1}\mathbb{E}\left(y_{n+1}-y_n\right)^4 \\
&= \mathcal{O}\left(\frac{\epsilon^{4+2\max(0,\alpha-2)}}{\Delta^2}\right) \\
&= \mathcal{O}\left(\frac{\epsilon^{\max(4,2\alpha)}}{\Delta^2}\right)
\end{aligned}$$

Adding up all terms for line (55), we have,

$$\sum_{i=1}^{3}\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_i^2\right)^2 = \mathcal{O}\left(\Delta^2+\epsilon^4+\frac{\epsilon^{\max(4,2\alpha)}}{\Delta^2}\right). \tag{58}$$

In line (56), for $i \in \{1, 2\}$, we have

$$\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_i\sqrt{\tilde{q}\Delta}\eta_n\right)^2 \leq CN^2\Delta\mathbb{E}\left(\hat{R}_i\eta_n\right)^2 = CN\mathbb{E}(\hat{R}_i)^2$$

Since $\mathbb{E}(\hat{R}_1)^2 = \mathcal{O}(\Delta^2)$, we have

$$\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_1\sqrt{\tilde{q}\Delta}\eta_n\right)^2 = \mathcal{O}(N\Delta^2) = \mathcal{O}(\Delta);$$

since $\mathbb{E}(\hat{R}_2)^2 = \mathcal{O}(\epsilon^2\Delta)$, we have

$$\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_2\sqrt{\tilde{q}\Delta}\eta_n\right)^2 = \mathcal{O}(N\epsilon^2\Delta) = \mathcal{O}(\epsilon^2).$$

Paper No. 15-05, www.warwick.ac.uk/go/crism

Again, it is different for $\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_3\sqrt{\tilde{q}\Delta}\eta_n\right)^2$ due to correlation between $\hat{R}_3^{(n)}$ and η_n. Using the expression from (43) by only considering the dominating terms, we have

$$\begin{aligned}
& \mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_3\sqrt{\tilde{q}\Delta}\eta_n\right)^2 \\
= & \ \mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_3^2\left(\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) \\
+ & \ \mathbb{E}\left(\frac{1}{T^2}\sum_{m\neq n}\hat{R}_3^{(m)}\hat{R}_3^{(n)}\int_{m\Delta}^{(m+1)\Delta}\sqrt{\tilde{q}}dWs\int_{n\Delta}^{(n+1)\Delta}\sqrt{\tilde{q}}dWs\right)
\end{aligned}$$

By computing the order of the dominating terms and the martingale terms, when $m = n$,

$$\begin{aligned}
\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_3^2\left(\sqrt{\tilde{q}\Delta}\eta_n\right)^2\right) &= \frac{1}{T}\sum_{n=0}^{N-1}\Delta\mathbb{E}\left(\hat{R}_3^2\tilde{q}\eta_n^2\right) \\
&= \frac{1}{T}\mathbb{E}(\hat{R}_3^2\eta_n^2) \\
&= \mathcal{O}\left(\epsilon^{\max(\alpha,2)}\right)
\end{aligned}$$

and when $m < n$,

$$\begin{aligned}
& \mathbb{E}\left(\frac{1}{T^2}\sum_{m\neq n}\hat{R}_3^{(m)}\hat{R}_3^{(n)}\int_{m\Delta}^{(m+1)\Delta}\sqrt{\tilde{q}}dWs\int_{n\Delta}^{(n+1)\Delta}\sqrt{\tilde{q}}dWs\right) \\
\leq & \ CN^2\epsilon^2\mathbb{E}\left((y_{n+1}-,y_n)(y_{m+1}-y_m)\right. \\
\times & \left.\int_{n\Delta}^{(n+1)\Delta}dW_s'\int_{m\Delta}^{(m+1)\Delta}dW_s\right) \\
\leq & \ CN^2\epsilon^2\mathbb{E}\left((y_{n+1}-y_n)\int_{n\Delta}^{(n+1)\Delta}dW_s'\right. \\
\times & \left.\mathbb{E}\left((y_{m+1}-y_m)\int_{m\Delta}^{(m+1)\Delta}dW_s|\mathcal{F}_{m\Delta}\right)\right)
\end{aligned}$$

Using the expansion in (43), and using the dominating terms only,

$$\begin{aligned}
& \mathbb{E}\left((y_{m+1}-y_m)\int_{n\Delta}^{(n+1)\Delta}dW_s|\mathcal{F}_{m\Delta}\right) \\
= & \ \mathbb{E}\left(\left((e^{-\frac{\Delta}{\epsilon^2}}-1)y_m\right.\right. \\
+ & \ \frac{1}{\epsilon^2}\int_{m\Delta}^{(m+1)\Delta}e^{-\frac{(m+1)\Delta-s}{\epsilon^2}}x(s)ds \\
+ & \left.\left.\frac{1}{\epsilon}\int_{m\Delta}^{(m+1)\Delta}e^{-\frac{(m+1)\Delta-s}{\epsilon^2}}dV_s\right)\int_{m\Delta}^{(m+1)\Delta}dW_s|\mathcal{F}_{m\Delta}\right) \\
= & \ \mathcal{O}(\epsilon(e^{-\frac{\Delta}{\epsilon^2}}-1))
\end{aligned}$$

Therefore, when $m < n$, we have,

$$\begin{aligned}
&\mathbb{E}\left(\frac{1}{T^2}\sum_{m\neq n}\hat{R}_3^{(m)}\hat{R}_3^{(n)}\int_{m\Delta}^{(m+1)\Delta}\sqrt{\tilde{q}}dWs\int_{n\Delta}^{(n+1)\Delta}\sqrt{\tilde{q}}dWs\right)\\
= \;& \mathcal{O}(\frac{\epsilon^4}{\Delta^2}(e^{-\frac{\Delta}{\epsilon^2}}-1)^2)\\
= \;& \mathcal{O}(\epsilon^{4-2\alpha+2\max(\alpha-2,0)})\\
= \;& \mathcal{O}(\epsilon^{\max(0,4-2\alpha)})
\end{aligned}$$

In the case $m > n$, the result is identical due to symmetry. Adding up all terms for line (56),

$$\begin{aligned}
&\sum_{i=1}^{5}\mathbb{E}\left(\frac{1}{T}\sum_{n=0}^{N-1}\hat{R}_i\sqrt{\tilde{q}\Delta}\eta_n\right)^2\\
&= \mathcal{O}\left(\Delta+\epsilon^2+\epsilon^{\max(\alpha,2)}+\epsilon^{2\max(0,2-\alpha)}\right)
\end{aligned} \tag{59}$$

In line (57), we have

$$\sum_{i\neq j}\mathbb{E}\left(\sum_{n=0}^{N-1}\hat{R}_i\hat{R}_j\right)^2 \leq N\mathbb{E}(R_i)^2\mathbb{E}(R_j)^2$$

Substituting in the L^2 norms of each $\hat{R}_i$, $i \in \{1,2,3\}$, we have for line (57),

$$\begin{aligned}
&\sum_{i\neq j}\mathbb{E}\left(\sum_{n=0}^{N-1}\hat{R}_i\hat{R}_j\right)^2\\
= \;& \mathcal{O}\left(\Delta^2\epsilon^2+\Delta\epsilon^{\max(\alpha,2)}+\epsilon^{2+\max(\alpha,2)}\right)
\end{aligned} \tag{60}$$

Aggregating bounds (58), (59) and (60) for equation lines from (54) to (57) respectively, we have

$$\begin{aligned}
&\mathbb{E}(\hat{q}_\epsilon-\tilde{q})^2\\
= \;& \mathcal{O}(\Delta)\\
+ \;& \mathcal{O}\left(\Delta^2+\epsilon^4+\frac{\epsilon^{\max(4,2\alpha)}}{\Delta^2}\right)\\
+ \;& \mathcal{O}\left(\Delta+\epsilon^2+\epsilon^{\max(\alpha,2)}+\epsilon^{2\max(0,2-\alpha)}\right)\\
+ \;& \left(\Delta^2\epsilon^2+\Delta\epsilon^{\max(\alpha,2)}+\epsilon^{2+\max(\alpha,2)}\right)
\end{aligned}$$

It is clear that when $\alpha < 2$,

$$\mathbb{E}(\hat{q}_\epsilon-\tilde{q})^2 = \mathcal{O}(\Delta+\epsilon^{4-2\alpha}+\epsilon^2).$$

The error is minimized when $\alpha = 4/3$, which is of order

$$\mathbb{E}(\hat{q}_\epsilon-\tilde{q})^2 = \mathcal{O}\left(\epsilon^{\frac{4}{3}}\right).$$

It is easy to see when $\alpha > 2$, the error explodes. This completes the proof. □

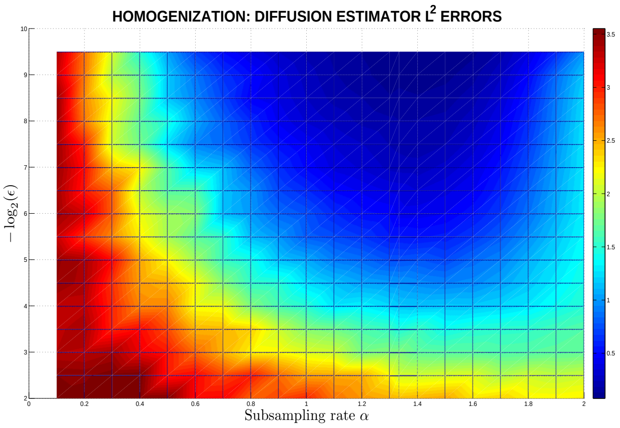

Figure 4: Homogenization: L^2 norm of $(\hat{q}_\epsilon - \tilde{q})$ for different ϵ and α

In Figure 4, we show an example of the L^2 error of the diffusion parameter with various scaling parameter ϵ and subsampling rate α. We see that the error is minimized around $\alpha = 4/3$ as in Theorem 3.8.

References

[1] Aït-Sahalia, Y.; Mykland, P. A; Zhang, L. A Tale of Two Time Scales: Determining Integrated Volatility With Noisy High-Frequency Data *Journal of the American Statistical Association* **100(472)** 2005

[2] Aït-Sahalia, Y.; Mykland, P. A; Zhang, L. How Often To Sample a Continuous-Time Process in the Presence of Market Microstructure Noise *The Review of Financial Studies* **18** 2005

[3] Azencott, R.; Beri A.; Timofeyev, I. Adaptive sub-sampling for parametric estimation of gaussian diffusions. *Journal of Statistical Physics* **139(6)**, 2010

[4] Azencott, R.; Beri A.; Timofeyev, I. Sub-sampling and parametric estimation for multiscale dynamics. *SIAM Multiscale Model. Simul.* 2010

[5] Anderson, D. F.; Cracium, G.; Kurtz, T. Product-form Stationary Distributions for Deficiency Zero Chemical Reaction Networks. *Bulletin of mathematical biology* **72(8)**, 2010

[6] Bensoussan, A.; Lions, J.-L.; Papanicolaou. G. . Asymptotic analysis for periodic structures, *Studies in Mathematics and its Applications* **5** 1978.

[7] Bishwal, J.P.N. Parameter Estimation in Stochastic Differential Equations Springer-Verlag 2008

[8] Chen, X. Limit theorems for functionals of ergodic Markov Chains with general state space. *Memoirs of the AMS* **129(664)**, 1999

[9] Cormen, T. H. *et al.* Introduction to Algorithms. (Second Edition) MIT Press and McGraw-Hill, 2001.

[10] Crisan, D. Stochastic Analysis Springer, 2010

[11] Givon, D. ; Kevrekidis, I.G.; Kupferman, R. Strong convergence schemes of projective intregration schemes for singularly perturbed stochastic differential equations. *Comm. Math. Sci.* **4(4)** 2006

[12] Graversen, S. E.; Peskir, G. Maximal Inequalities for the Ornstein-Uhlenbeck Process *Proceedings of the American Mathematical Society* **128(10)**, 2000

[13] Hall, P. and Heyde, C.C Martingale limit theory and its application. Academic Press, 1980.

[14] Isaacson, E.; Keller, H. B. Analysis of Numerical Methods Wiley & Sons. 1966.

[15] Kallenberg, O. Foundations of Modern Probability (Second Edition) Springer, 2002.

[16] Koval, V. A. On the Strong Law of Large Numbers for Multivariate Martingales with Continuous Time. *Ukranian Mathematical Journal* **53(9)** 2001

[17] Kutoyants, Y. A. Statistical inference for ergodic diffusion processes. Springer-Verlag, 2004.

[18] Liu, D.; Strong convergence of principle of averaging for multiscale stochastic dynamcial systems *Comm. Math. Sci* **8(4)** 2010

[19] Liu, W. E, D.; Vanden-Eijnden, E. Analysis of multiscale methods for stochastic differential equations. *Comm. Pure Appl. Math.* **58(11)** 2005

[20] Miao, W. Estimation of diffusion parameters in diffusion processes and their asymptotic normality *International Journal of Comtemporary Mathematical Science* **1(16)** 2006

[21] Olhede, S; Pavliotis, G. A. ; Sykulski, A. Multiscale Inference for High Frequency Data *Tech. Rep. 290 Department of Statistical Science, UCL* 2008

[22] Papanicolaou, G C., Stroock, D W., Varadhan, S R S. Martingale approach to some limit theorems. In Papers from the Duke Turbulence Conference *Paper No. 6, Duke Univ. Math. Ser.*, **Vol. III.** 1977

[23] Papavasiliou, A.; Pavlioitis, G. A.; Stuart, A. M. Maximum Likelihood Drift Estimation for Multiscale Diffusions *Stoch. Proc. Appl.* **119(10)** 2009

[24] Pavliotis, G.A.; Stuart, A.M. White Noise Limits for Intertial Particles in a Random Field. *SIAM J. Multiscale Modeling and Simulation* **1(4)** 2003

[25] Pavliotis, G. A. ; Stuart, A. M. Multiscale Methods Averaging and Homogenization. Springer, 2008.

[26] Pavliotis, G. A. ; Stuart, A. M. Parameter Estimation for Multiscale Diffusions *J. Stat. Phys.* **127** 2007

[27] Prakasa Rao, B. L. S. Semimartingales and their statistical inference. *Monograph on Statistics and Applied Probability*, **83** Chapman & Hall/CRC, 1999

[28] Vanden-Eijnden, E. Numerical techniques for multi-scale dynamical systems with stochastic effects. *Commun. Math. Sci.* **1(2)** 2003