SAMPLE PARTITIONING ESTIMATION FOR ERGODIC DIFFUSIONS

LUÍS P. RAMOS, PEDRO P. MOTA, AND JOÃO T. MEXIA

Abstract. In this paper we present a new technique to obtain estimators for parameters of ergodic processes. When a diffusion is ergodic its transition density converges to the invariant density [1]. This convergence enabled us to introduce a sample partitioning technique that gives, in each sub-sample, observations that can be treated as independent and identically distributed. Within this framework, is possible the construction of estimators like maximum likelihood estimators or others.

Consistency; Ergodic Diffusions; Independency; Least Squares; Martingale Estimating Functions; Maximum Likelihood Estimators; Method of Moments; Transition and Invariant Densities.

1. Introduction

Diffusion processes are widely used in different areas as in biology, economics and financial mathematics [6] or [10], among many others. Typicaly, the problems are stated for independent and identically distributed observations but, when the processes are ergodic we do not need this hipothesis. For ergodic processes many estimation techniques are well studied [5], some of those techniques use the transition density of the process but, it turns out that in many cases this transition density is untractable and cannot be used for estimation purposes. Using the ergodic nature of the processes and the invariant density we propose a sample partitioning technique that gives, in each subsample, observations that can be treated as independent and identically distributed. In fact, for the Ornstein-Uhlenbeck process we prove in proposition 4.2 that any two observations are asymptotically independent when the time lag between the two increases. Within this framework, is possible the construction of estimators, in each subsample, like maximum likelihood estimators or others, using the invariant density. The average of the subsamples estimators is a natural choice as estimators for the process parameters. In the case of the Ornstein-Uhlenbeck and for a family of processes that includes the Cox-Ingersoll & Ross and the Dixit & Pindyck process, we prove the consistency of the sample partitioning estimators. For the Ornstein-Uhlenbeck and the Cox-Ingersoll & Ross processes we compare the results obtained from our estimators with the results obtained from the usual estimators for those processes, see [8]. In a simulation study we show that our estimators will have smaller MSE than the usual estimators and in some cases the diference will be very large. For the Dixit & Pindyck and other processes for which there are no explicit estimators we only present our estimation technique results.

The remainder of this article is organized as follows. In section 2 we introduce the ergodic diffusions that we will study and the sub-sampling technique is described in section 3. In section 4 we present the application of the technique to the ergodic processes presented in section 2, we build the sample partitioning estimators, prove their consistency and for the Ornstein-Uhlenbeck process we will prove in proposition 4.2 a fundamental result about asymptotic independency. In section 4 we, also, implement the simulation study where we will compare our estimators with the usual ones for the Ornstein-Uhlenbeck and Cox-Ingersoll & Ross processes by comparing the MSE. We will compute the parameter estimators for other ergodic processes and, finally, in section 5 we will give some conclusions.

2. Diffusions and Ergodicity

In what follows we will work with diffusion processes, see [2] or [9], that is, a time-homogeneous stochastic process satisfying the stochastic differential equation (SDE),

$$dX_t = b(X_t, \boldsymbol{\theta})dt + \sigma(X_t, \boldsymbol{\theta})dB_t, \tag{2.1}$$

where $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p$ is a multidimensional parameter and $\{B_t\}_{t\geq 0}$ is the Brownian motion or Wiener process. The functions $b : \mathbb{R} \times \Theta \to \mathbb{R}$ and $\sigma : \mathbb{R} \times \Theta \to]0, +\infty[$ are known and such that the solution of (2.1) exists, see [3] or [7].

The function b is known as the drift coefficient of X, the function σ is known as the diffusion coefficient of X.
An example of a diffusion is the Ornstein-Uhlenbeck process, that is, the process solution of the equation,

$$dX_t = b(a - X_t)dt + \sigma dB_t, b > 0.$$

We will be interested in the study of a special type of diffusions, the ergodic diffusions, for a review in statistical inference for diffusion ergodic processes see [5]. It is well known the following characterization of an ergodic process. With $E =]l, r[$ the range of variation of X_t, $t > 0$, and $l < x_0 < r$, let the scale function and speed measure, be

$$\begin{cases} s(x, \boldsymbol{\theta}) = \exp\left[-2\int_{x_0}^{x} \frac{b(y,\boldsymbol{\theta})}{\sigma^2(y,\boldsymbol{\theta})}dy\right] \;\; ; \; x \in]l; r[\\ \\ m(x, \boldsymbol{\theta}) = \frac{1}{s(x,\boldsymbol{\theta})\sigma^2(x,\boldsymbol{\theta})} \; ; \; x \in]l; r[\end{cases} \tag{2.2}$$

respectively.

The diffusion is ergodic whenever, for every $\boldsymbol{\theta} \in \Theta$,

$$\int_{x_0}^{r} s(x, \boldsymbol{\theta})dx = \int_{l}^{x_0} s(x, \boldsymbol{\theta})dx = +\infty \tag{2.3}$$

and

$$M(\boldsymbol{\theta}) = \int_{l}^{r} m(x, \boldsymbol{\theta})dx < +\infty. \tag{2.4}$$

With the invariant density being,

$$f_{\boldsymbol{\theta}}(x) = \frac{m(x, \boldsymbol{\theta})}{M(\boldsymbol{\theta})} \;\; ; \; x \in]l; r[. \tag{2.5}$$

We have, regarding the invariant density, that:

Theorem 2.1. *The transition density of a ergodic diffusion tends to the corresponding invariant density when the time lag tends to infinite* [1].

Next we will introduce some ergodic processes that we will study in this paper.

Example 1. **Ornstein-Uhlenbeck Process**

$$dX_t = b(a - X_t)dt + \sigma dB_t, b > 0. \tag{2.6}$$

It is known that the process, solution os this stochastic differential equation is ergodic, with invariant density,

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{\sqrt{2\pi \frac{\sigma^2}{2b}}} e^{-\frac{(x-a)^2}{2\left(\frac{\sigma^2}{2b}\right)}} \sim N\left(a; \frac{\sigma^2}{2b}\right).$$

Example 2. **A special class of mean reverting processes**

The processes, solution of the SDE

$$dX_t = b(a - X_t)X_t^\gamma dt + \sigma\sqrt{X_t^{\gamma+1}}dB_t, a, b > 0, \gamma \geq 0, \tag{2.7}$$

are ergodic, when $2ab > \sigma^2(\gamma + 1)$, because, with $\boldsymbol{\theta} = (a, b, \gamma, \sigma)$,

$$\begin{aligned} S(x; \boldsymbol{\theta}) &= \int_{x_0}^{x} s(y, \boldsymbol{\theta})dy = \int_{x_0}^{x} \exp\left(-2\int_{x_0}^{y} \frac{b(a-v)v^\gamma}{\sigma^2 v^{\gamma+1}}dv\right)dy \\ &= x_0^{\frac{2ab}{\sigma^2}} e^{-\frac{2b}{\sigma^2}x_0} \int_{x_0}^{x} y^{-\frac{2ab}{\sigma^2}} e^{\frac{2b}{\sigma^2}y} dy \to +\infty, x \to +\infty, x \to 0, \end{aligned}$$

and

$$\begin{aligned} M(\boldsymbol{\theta}) &= \int_0^{+\infty} \frac{1}{\sigma^2 x^{\gamma+1}} \exp\left(2\int_{x_0}^{x} \frac{b(a-v)v^\gamma}{\sigma^2 v^{\gamma+1}}dv\right)dx = \\ &= \frac{x_0^{-\frac{2ab}{\sigma^2}} e^{\frac{2b}{\sigma^2}x_0}}{\sigma^2} \int_0^{\infty} x^{\frac{2ab}{\sigma^2}-\gamma-1} e^{-\frac{2b}{\sigma^2}x} dx < \infty, \text{ if } 2ab > \sigma^2(\gamma+1), \end{aligned}$$

having invariant density,

$$\begin{aligned} f_{\boldsymbol{\theta}}(x) &= \frac{x_0^{-\frac{2ab}{\sigma^2}} e^{\frac{2b}{\sigma^2}x_0}}{\sigma^2} x^{\frac{2ab}{\sigma^2}-\gamma-1} e^{-\frac{2b}{\sigma^2}x} \left(\frac{x_0^{-\frac{2ab}{\sigma^2}} e^{\frac{2b}{\sigma^2}x_0}}{\sigma^2} \int_0^{\infty} x^{\frac{2ab}{\sigma^2}-\gamma-1} e^{-\frac{2b}{\sigma^2}x} dx\right)^{-1} \\ &= \frac{x^{\alpha-1}e^{-\beta x}\beta^\alpha}{\Gamma(\alpha)} \sim Gamma(\alpha, \beta), \text{ with, } \alpha = \frac{2ab}{\sigma^2} - \gamma = \frac{2ab - \sigma^2\gamma}{\sigma^2}, \beta = \frac{2b}{\sigma^2}. \end{aligned}$$

Remark 2.2. Within this class of processes, we have,

Cox-Ingersoll & Ross ($\gamma = 0$)

$$dX_t = b(a - X_t)dt + \sigma\sqrt{X_t}dB_t;$$

Dixit & Pindyck process ($\gamma = 1$)

$$dX_t = b(a - X_t)X_t dt + \sigma X_t dB_t.$$

Example 3. **An ergodic process with Cauchy invariant density**

The solution of,

$$dX_t = \frac{-\sigma^2(X_t - \alpha)}{1 + (X_t - \alpha)^2}dt + \sigma dB_t. \tag{2.8}$$

is an ergodic process, because it verifies, with $\boldsymbol{\theta} = (\alpha, \sigma)$,

$$S(x;\boldsymbol{\theta}) = \int_{x_0}^{x} \exp\left(-2\int_{x_0}^{y} \frac{-\sigma^2(v-\alpha)}{\sigma^2(1+(v-\alpha)^2)}dv\right)dy$$
$$= \frac{1}{1+(x_0-\alpha)^2}\int_{x_0}^{x} 1+(y-\alpha)^2 dy \to +\infty, x \to \pm\infty$$

and

$$M(\boldsymbol{\theta}) = \int_{-\infty}^{+\infty} \frac{1}{\sigma^2}\exp\left(2\int_{x_0}^{x} \frac{-\sigma^2(v-\alpha)}{\sigma^2(1+(v-\alpha)^2)}dv\right)dx =$$
$$= \frac{1+(x_0-\alpha)^2}{\sigma^2}\pi$$

having invariant density,

$$f_{\boldsymbol{\theta}}(x) = \frac{1+(x_0-\alpha)^2}{\sigma^2(1+(x-\alpha)^2)}\left(\frac{1+(x_0-\alpha)^2}{\sigma^2}\pi\right)^{-1}$$
$$= \frac{1}{\pi\left(1+(x-\alpha)^2\right)} \sim Cauchy(\alpha, 1).$$

3. Limit Independence and Sub-Sampling

Suppose that we have discrete observations, $X_1, ..., X_n$, of the process at times $t_i = i\Delta$ with $\Delta = t_{i+1} - t_i, i = 1, ..., n-1$, that is, equally spaced observations of the process. The main idea in our estimation technique is that, for ergodic processes, if we consider two observations separated by a time interval long enough then they can be treated as if they are independent and both with the same invariant density. When the transition density converges to the invariant density we assume that there is $\overline{\Delta}$ such that, for, $t > \overline{\Delta}$, we may assume X_t to have the invariant density and, when $\overline{\Delta} < t < t' - \overline{\Delta}$, X_t and $X_{t'}$ can be treated as if they are independent. The best way to find such a $\overline{\Delta}$ is an open problem, but for the Ornstein-Uhlenbeck process we have a result based in the correlation coefficient, as we will see in proposition 4.2, that should be enough to the study of $\overline{\Delta}$. Another possibility to find $\overline{\Delta}$ should be through hypothesis testing. Assuming that we have $m, k \in \{1, ..., n\}$ such that $m\Delta > \overline{\Delta}$ and $n = km$, we put the observations X_j at times t_j, $j = 1, ..., km$, in the following table:

$$\begin{array}{ccccc}
X_1 & \cdots & X_j & \cdots & X_m \\
\vdots & & \vdots & & \vdots \\
X_{(i-1)m+1} & \cdots & X_{(i-1)m+j} & \cdots & X_{im} \\
\vdots & & \vdots & & \vdots \\
X_{(k-1)m+1} & \cdots & X_{(k-1)m+j} & \cdots & X_{km}
\end{array}$$

Now, the matched sub-samples corresponding to the lines of the table will have the same distribution and we can treat the observations in each column as being independent and identically distributed with the invariant density. Thus, from each column we can obtain a estimator (maximum likelihood estimator if possible, or other kind of estimator) $\boldsymbol{\theta}_j,\ j = 1, \ldots, m$ for $\boldsymbol{\theta}$. Afterwards, we take the average of the $\boldsymbol{\theta}_j,\ j = 1, \ldots, m$, in order to estimate $\boldsymbol{\theta}$.

4. Application to ergodic processes

In this section we will apply the sub-sampling estimation method for some of the ergodic processes presented in section 2.

Application 1. Example 1 (Ornstein-Uhlenbeck process).
Consider,

$$dX_t = b(a - X_t)dt + \sigma dB_t, b > 0 \tag{4.1}$$

the solution of this SDE is,

$$X_t = X_0 e^{-bt} + a(1 - e^{-bt}) + \sigma \int_0^t e^{-b(t-s)} dB_s$$

being the conditional mean of X_{i+1} given X_i,

$$\mathbb{E}[X_{i+1}|X_i] = X_i e^{-b\Delta} + a(1 - e^{-b\Delta}),$$

the conditional variance,

$$\mathbb{V}[X_{i+1}|X_i] = \frac{\sigma^2}{2b}\left(1 - e^{-2b\Delta}\right),$$

and the conditional distribution of X_{i+1} given X_i is then

$$N\left(X_i e^{-b\Delta} + a(1 - e^{-b\Delta}); \frac{\sigma^2}{2b}\left(1 - e^{-2b\Delta}\right)\right).$$

As we saw before, the Ornstein-Uhlenbeck process is ergodic with invariant density,

$$f_{\boldsymbol{\theta}}(x) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{2b}}} e^{-\frac{(x-a)^2}{2\left(\frac{\sigma^2}{2b}\right)}} \sim N\left(a; \frac{\sigma^2}{2b}\right).$$

If X_0 is a random variable with the invariant density then the process is strictly stationary and then for $\forall t \geq 0, X_t$ have the $N\left(a; \frac{\sigma^2}{2b}\right)$ distribution. For simplicity we will suppose that σ is known and we will estimate only a and b.

From the sub-sampling approach, we obtained the maximum likelihood estimators for a and b,

$$\widehat{a}_n = \frac{1}{m}\sum_{j=1}^{m} \widehat{a}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{k}\sum_{i=1}^{k} X_{(i-1)m+j}\right] \tag{4.2}$$

and

$$\widehat{b}_n = \frac{1}{m}\sum_{j=1}^{m} \widehat{b}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{2}{k}\sum_{i=1}^{k}\left(X_{(i-1)m+j} - \widehat{a}_{j,k}\right)^2\right]^{-1} \tag{4.3}$$

We have a first result on consistency.

Proposition 4.1. *The sample partitioning, maximum likelihood estimators of a and b in the Ornstein-Uhlenbeck process, are consistent.*

Proof. Let ξ be a random variable with the invariant density, that is, with a normal density of mean a and variance $\frac{\sigma^2}{2b}$, $b > 0$. For $\widehat{a}_n$, we have,

$$\widehat{a}_n = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{k}\sum_{i=1}^{k} X_{(i-1)m+j}\right] = \frac{1}{n}\sum_{i=1}^{n} X_i \overset{n\to\infty}{\longrightarrow} \mathbb{E}[\xi] = a \quad (a.s.) \tag{4.4}$$

where we only need to use the ergodic theorem [5].
For $\widehat{b}_n$, notice that, with m fixed, $n \to \infty \Rightarrow k = k(n) \to \infty$ and to prove that

$$\widehat{b}_n = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{2}{k}\sum_{i=1}^{k}\left(X_{(i-1)m+j} - \widehat{a}_{j,k}\right)^2\right]^{-1} \overset{n\to\infty}{\longrightarrow} \frac{b}{\sigma^2} \tag{4.5}$$

is enough to prove that

$$\frac{2}{k}\sum_{i=1}^{k}\left(X_{(i-1)m+j} - \widehat{a}_{j,k}\right)^2 \overset{n\to\infty}{\longrightarrow} \frac{\sigma^2}{b}, \forall m$$

but

$$\frac{1}{k}\sum_{i=1}^{k}\left(X_{(i-1)m+j} - \widehat{a}_{j,k}\right)^2 = \frac{1}{k}\sum_{i=1}^{k}\left(X_{(i-1)m+j} - a\right)^2 - \left(\widehat{a}_{j,k} - a\right)^2 \overset{n\to\infty}{\longrightarrow} \frac{\sigma^2}{2b} \quad (a.s.)$$

because from the ergodic theorem the first term converges to $\mathbb{V}[\xi]$ and the second to zero. □

We have a result about the asymptotic independency between the variables in each subsample when the process is the Ornstein-Uhlenbeck process.

Proposition 4.2. *Let $X_1, ..., X_n, ...$ be observations of the Ornstein-Uhlenbeck process, let X and Y be random variables, such that, $X \in \{X_1, ..., X_p\}$ and $Y \in \{X_{p+m}, ...\}$, for some m. Then the correlation coefficient,*

$$\rho_m(X, Y) \to 0, \ \text{as } m \to \infty \tag{4.6}$$

exponentially fast.

Proof. Put $X = X_i$ for some $i \in \{1, ..., p\}$ and $Y = X_j$ for some $j \geq p + m$. Then we have from the ergodicity and the invariant density that,

$$\mathbb{E}[X_i] = \mathbb{E}[X_j] = a$$

and from the conditional distribution, Markov property and conditional expectation properties, with $\mathcal{F}_i = \sigma\{X_1, ..., X_i\}$, the $\sigma-$field generated by $X_1, ..., X_i$,

$$\mathbb{E}[X_iX_j] = \mathbb{E}[\mathbb{E}[X_iX_j|\mathcal{F}_i]] = \mathbb{E}[X_i\mathbb{E}[X_j|\mathcal{F}_i]] = \mathbb{E}\left[X_i\left(X_ie^{-b(j-i)\Delta} + a(1 - e^{-b(j-i)\Delta})\right)\right]$$
$$= \mathbb{E}\left[X_i^2e^{-b(j-i)\Delta}\right] + \mathbb{E}\left[X_ia(1 - e^{-b(j-i)\Delta})\right] = e^{-b(j-i)\Delta}\mathbb{E}\left[X_i^2\right] + a(1 - e^{-b(j-i)\Delta})\mathbb{E}\left[X_i\right]$$
$$= e^{-b(j-i)\Delta}\left(\frac{\sigma^2}{2b} + a^2\right) + a^2(1 - e^{-b(j-i)\Delta}) = \frac{\sigma^2}{2b}e^{-b(j-i)\Delta} + a^2$$

because

$$\mathbb{V}[X_i] = \mathbb{V}[X_j] = \frac{\sigma^2}{2b}$$

becoming in this case,

$$\frac{|\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]|}{\sqrt{\mathbb{V}[X_i]\mathbb{V}[X_j]}}=\frac{\left|\frac{\sigma^2}{2b}e^{-b(j-i)\Delta}+a^2-a^2\right|}{\frac{\sigma^2}{2b}}=e^{-b(j-i)\Delta}. \tag{4.7}$$

Then we get a measure of the dependency between any two variables $X_i, i=1,...,p$ and $X_j, j\geq p+m$ given by,

$$\rho^*(m)=\sup_{i,j}\left\{\frac{|\mathbb{E}[X_iX_j]-\mathbb{E}[X_i]\mathbb{E}[X_j]|}{\sqrt{\mathbb{V}[X_i]\mathbb{V}[X_j]}}:i=1,...,p;j\geq p+m\right\}=e^{-bm\Delta} \tag{4.8}$$

and the result follows because $\rho_m\leq\rho^*(m)$. □

We now compare our estimators with the well known estimators for a and b,

$$\tilde{a}_n=\frac{S_yS_{xx}-S_{xy}S_x}{n(S_{xx}-S_{xy})+S_x(S_y-S_x)}$$

and

$$\tilde{b}_n=-\frac{1}{\Delta}\ln\left[\frac{S_{xy}-\tilde{a}_n(S_x+S_y)+\tilde{a}_n^2n}{S_{xx}-2\tilde{a}_nS_x+\tilde{a}_n^2n}\right]$$

when the argument of the logarithm is positive and where

$$S_x=\sum_{i=0}^{n-1}X_i;S_y=\sum_{i=0}^{n-1}X_{i+1};S_{xx}=\sum_{i=0}^{n-1}X_i^2;S_{yy}=\sum_{i=0}^{n-1}X_{i+1}^2;S_{xy}=\sum_{i=0}^{n-1}X_iX_{i+1}.$$

Remark 4.3. Notice that, a first advantage of our technique is that, for the Ornstein-Uhlenbeck, the estimator $\widehat{b}_n$ is always well defined while $\tilde{b}_n$ is well defined only when the argument in the logarithm function is positive.

We used the transition density, which is known, to simulate trajectories of the Ornstein-Uhlenbeck diffusion considering different values of b and for a fixed value of a. We obtained the results presented in the following tables with mean, standard deviation and the mean square error values for the estimators under analysis. We have fixed $\sigma=1$ and the time lag, $\Delta=1$, the results are from 1000 observations in 500 simulated trajectories.

Table 1. Mean, standard deviation and mean square error for both estimators of b when $k=100$, $m=10$

b	Mean($\widehat{b}_n$)	Mean($\tilde{b}_n$)	SD($\widehat{b}_n$)	SD($\tilde{b}_n$)	MSE($\widehat{b}_n$)	MSE($\tilde{b}_n$)
0.01	0.015	0.015	0.007	0.007	7E-5	7E-5
0.1	0.106	0.105	0.017	0.018	3E-4	3E-4
0.5	0.517	0.508	0.035	0.044	1.5E-3	1.9E-3
1	1.034	1.015	0.055	0.081	0.004	0.007
2	2.065	2.062	0.097	0.255	0.014	0.069
5	5.159	4.159	0.238	1.229	0.082	2.213
10	10.317	4.188	0.477	1.175	0.327	35.153

Looking at Table 1, we can see that the results for our estimator of b are better than least squares estimators when we consider a decomposition ($k=100$, $m=10$). In Figure 1 we plot the empirical versus the theoretical correlation coefficient and we can see that they are very close to each other being that consistent with the

TABLE 2. Mean, standard deviation and mean square error for both estimators of $a(a = 0)$ when $k = 100$, $m = 10$ and different values of b

b	Mean($\widehat{a}_n$)	Mean($\widetilde{a}_n$)	SD($\widehat{a}_n$)	SD($\widetilde{a}_n$)	MSE($\widehat{a}_n$)	MSE($\widetilde{a}_n$)
0.01	0.104	0.157	2.875	3.086	8.262	9.527
0.1	0.012	0.012	0.306	0.308	0.093	0.095
0.5	0.002	0.002	0.062	0.062	0.004	0.004
1	0.001	0.001	0.032	0.032	0.001	0.001
2	7E-4	7E-4	0.018	0.018	3E-4	3E-4
5	4E-4	4E-4	0.010	0.010	9E-5	9E-5
10	3E-4	3E-4	0.007	0.007	5E-5	5E-5

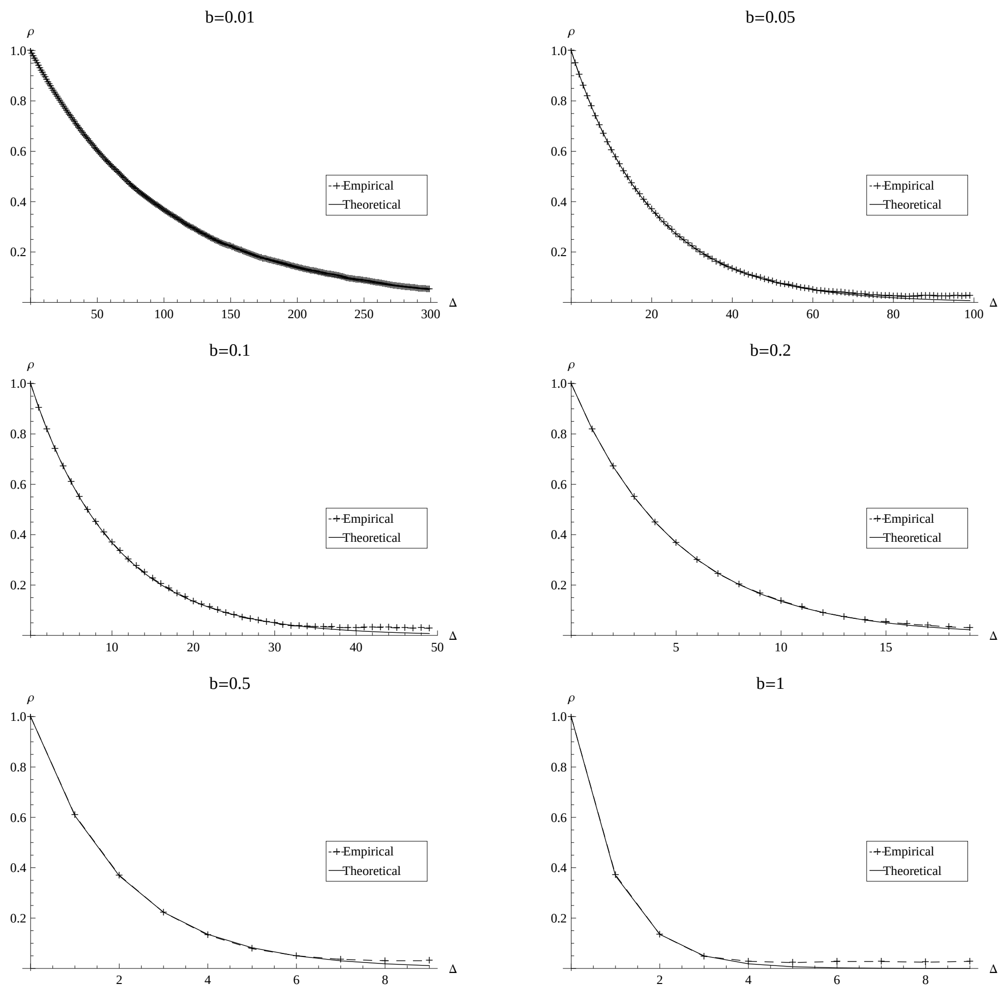

FIGURE 1. Empirical vs theoretical correlation coefficient (ρ)

proposition 4.2. The rate of convergence of the transition density to the invariant density for the Ornstein-Uhlenbeck diffusion depends on b and is faster for larger values of b, so our results are better for these values of b.

Since the parameter a does not change the process dynamics but only the mean reverting position we only present results for $a = 0$. For this parameter there is no difference between the two estimators for different values of b, as we can see in Table 2.

Application 2. Now we will apply our estimation procedure for the class of processes presented in example 2, for different values of γ and with the $ab > \frac{\sigma^2(\gamma+1)}{2}$ condition satisfied. First we will look at the Cox-Ingersoll & Ross process ($\gamma = 0$) and then we will get assymptotic results on consistency of the parameters estimators for the general case.

The invariant density for the Cox-Ingersoll & Ross process is the gamma density, $Gamma(\alpha, \beta)$, with $\alpha = \frac{2ab}{\sigma^2}, \beta = \frac{2b}{\sigma^2}$. For simplicity we take $\sigma = 1$ and deduce the estimators for a and b using in each subsample the maximum likelihood estimators for a and b.

We obtained,

$$\widehat{a}_n = \frac{1}{m}\sum_{j=1}^{m}\widehat{a}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{k}\sum_{i=1}^{k}X_{(i-1)m+j}\right] \tag{4.9}$$

and

$$\widehat{b}_n = \frac{1}{m}\sum_{j=1}^{m}\widehat{b}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\rho_j\left[\frac{2}{k}\sum_{i=1}^{k}X_{(i-1)m+j}\right]^{-1} \tag{4.10}$$

where, for $j = 1, ..., m$, ρ_j is solution of

$$\ln(\rho_j) - \psi(\rho_j) = \ln\left(\frac{1}{k}\sum_{i=1}^{k}X_{(i-1)m+j}\right) - \frac{1}{k}\sum_{i=1}^{k}\ln\left(X_{(i-1)m+j}\right)$$

with, $\psi(\rho_j)$ the digamma function. This equation must be solved using numerical methods.

Alternatively, we can deduce the estimators for a and b using in each subsample the moments estimators for the gamma density. We get in this case,

$$\breve{a}_n = \widehat{a}_n = \frac{1}{m}\sum_{j=1}^{m}\widehat{a}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\left[\frac{1}{k}\sum_{i=1}^{k}X_{(i-1)m+j}\right] \tag{4.11}$$

and

$$\breve{b}_n = \frac{1}{m}\sum_{j=1}^{m}\breve{b}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\widehat{a}_{j,k}\left[\frac{2}{k}\sum_{i=1}^{k}\left(X_{(i-1)m+j} - \widehat{a}_{j,k}\right)^2\right]^{-1}. \tag{4.12}$$

We now compare our estimators with the martingale estimating function estimators for a and b,

$$\tilde{b}_n = -\frac{1}{\Delta}\log\left(\frac{\sum_{i=0}^{n-1}X_{i+1}\sum_{i=0}^{n-1}X_i^{-1} - n\sum_{i=0}^{n-1}X_{i+1}X_i^{-1}}{\sum_{i=0}^{n-1}X_i\sum_{i=0}^{n-1}X_i^{-1} - n^2}\right)$$

if the logarithm argument is positive and

$$\tilde{a}_n = \frac{\sum_{i=0}^{n-1} X_{i+1}X_i^{-1} - ne^{-\tilde{b}_n\Delta}}{(1-e^{-\tilde{b}_n\Delta})\sum_{i=0}^{n-1} X_i^{-1}}.$$

We used the Euler scheme with a discretization step $\delta = 0.1$ to simulate trajectories of the Cox-Ingersoll & Ross process but we only work with the observations $X_0, X_1, X_2, ...,$ at times $0, 1, 2, ...,$ respectively. Again, we consider different values for b and we obtained the results presented in the following tables when comparing the sub-sampling estimators derived from maximum likelihood and moments estimators in the sub-samples with the traditional linear martingale estimating functions estimators. We have fixed $\sigma = 1$ and the time lag, $\Delta = 1$ and we consider $k \times m$ observations (≈ 1000) from 500 simulations.

Table 3. Mean, standard deviation and mean square error for both estimators of b when we use maximum likelihood estimators in the sub-samples

$k \times m$	b	Mean($\widehat{b}_n$)	Mean($\widetilde{b}_n$)	SD($\widehat{b}_n$)	SD($\widetilde{b}_n$)	MSE($\widehat{b}_n$)	MSE($\widetilde{b}_n$)
100×10	0.01	0.015	0.013	0.006	0.006	6E-5	4E-5
100×10	0.1	0.105	0.086	0.016	0.014	3E-4	4E-4
100×10	0.5	0.502	0.475	0.038	0.053	0.001	0.003
67×15	1	0.995	1.384	0.058	0.252	0.003	0.211
28×36	2	2.016	5.108	0.108	1.130	0.012	10.930
15×67	5	4.681	0.410	0.284	0.103	0.182	29.274
10×100	10	7.126	0.004	0.470	0.055	8.479	100.084

Table 4. Mean, standard deviation and mean square error for both estimators of b when we use moments estimators in the sub-samples

$k \times m$	b	Mean($\breve{b}_n$)	Mean($\widetilde{b}_n$)	SD($\breve{b}_n$)	SD($\widetilde{b}_n$)	MSE($\breve{b}_n$)	MSE($\widetilde{b}_n$)
67×15	0.01	0.015	0.013	0.007	0.006	7E-5	4E-5
67×15	0.1	0.101	0.086	0.016	0.014	3E-4	4E-4
50×20	0.5	0.500	0.475	0.039	0.053	1.5E-3	3.4E-3
40×25	1	0.993	1.384	0.058	0.252	0.003	0.211
20×50	2	2.008	5.108	0.108	1.130	0.012	10.930
10×100	5	4.824	0.410	0.314	0.103	0.130	29.274
10×100	10	6.417	0.004	0.423	0.055	13.014	100.084

We can observe from Tables 3 and 4 that the estimators obtained using maximum likelihood estimation or moment estimation in the sub-sampling do not produce very different results. As we can see the results for our estimators are similar to martingale estimating functions estimators when we consider optimal decompositions for $k \times m$ and small values of b, but are better when b increases.

For the general case, that is, for arbitrary γ, (but satisfying $ab > \frac{\sigma^2(\gamma+1)}{2}$), the invariant density, for the process solution of the equation in example 2, is the gamma density, $Gamma(\alpha, \beta)$, with $\alpha = \frac{2ab}{\sigma^2} - \gamma, \beta = \frac{2b}{\sigma^2}$. We will suppose that

γ is know and we deduce the estimators for a and b using in each subsample the moments estimators for the gamma density. We will get in this case, the equations,

$$\begin{cases} \frac{\alpha}{\beta} & = \bar{X}_j \\ \frac{\alpha}{\beta^2} & = S_j^2 \end{cases}, \tag{4.13}$$

with $\bar{X}_j$ the j^{th}-subsample mean and S_j^2 the j^{th}-subsample variance, for $j = 1, ..., m$.

Then we get the parameter estimators,

$$\breve{a}_n = \frac{1}{m}\sum_{j=1}^{m}\left[\bar{X}_j + \frac{S_j^2}{\bar{X}_j}\gamma\right] \tag{4.14}$$

and

$$\breve{b}_n = \frac{1}{m}\sum_{j=1}^{m}\breve{b}_{j,k} = \frac{1}{m}\sum_{j=1}^{m}\frac{\sigma^2\bar{X}_j}{2S_j^2}. \tag{4.15}$$

We have also a result on consistency for this estimators.

Proposition 4.4. *The sample partitioning, moments estimators of a and b in this class of processes, are consistent.*

Proof. Consider ξ the random variable with the invariant gamma density, $Gamma(\alpha, \beta)$, where $\alpha = \frac{2ab-\gamma\sigma^2}{\sigma^2}, \beta = \frac{2b}{\sigma^2}$. Is straightforward to prove the consistency of both estimators, since, using the ergodic theorem,

$$\bar{X}_j = \frac{1}{k}\sum_{i=1}^{k}X_{(i-1)m+j} \overset{n\to\infty}{\longrightarrow} \mathbb{E}[\xi] = \frac{\alpha}{\beta} \quad a.s. \tag{4.16}$$

and

$$S_j^2 = \frac{1}{k}\sum_{i=1}^{k}(X_{(i-1)m+j} - \bar{X}_j)^2 \overset{n\to\infty}{\longrightarrow} \mathbb{V}[\xi] = \frac{\alpha}{\beta^2} \quad a.s. \tag{4.17}$$

Then,

$$\bar{X}_j + \frac{S_j^2}{\bar{X}_j}\gamma \overset{n\to\infty}{\longrightarrow} \frac{\alpha}{\beta} + \frac{\gamma}{\beta} = a \quad a.s. \tag{4.18}$$

and

$$\frac{\sigma^2\bar{X}_j}{2S_j^2} \overset{n\to\infty}{\longrightarrow} \frac{\beta\sigma^2}{2} = b \quad a.s. \tag{4.19}$$

proving the consistency of the estimators in the sub-samples. Being the sample partitioning estimators just the mean of this estimators the result follows. □

When the γ parameter is equal to one, the process solution of the stochastic differential equation (2.7) is known as the Dixit & Pindyck process, for this process and the following others we do not have an explicit transition density and because of that our estimation method is more usefull than in the previous cases. As to the best of our knowledge there are no explicit estimators for the parameters of the next processes and that is the reason why we will not compare the estimates of the parameters. However, we will show, through simulation, that from the subsampling procedure we can get good results for the parameter estimators. In the simulation study we will use the Euler scheme to simulate the trajectories with a discretization step of $\delta = 0.01$ but keeping only the observations at integer times, that is, at times $i\Delta$, with $\Delta = 1$. For the Dixit & Pindyck process we get the results presented in Table 5.

TABLE 5. Mean and standard deviation for the estimate of b in the Dixit & Pindyck process ($\gamma = 1$)

$k \times m$	b	Mean($\breve{b}_n$)	SD($\breve{b}_n$)
100×10	0.01	0.0100	0.0009
67×15	0.1	0.1005	0.0060
67×15	0.5	0.4955	0.0251
36×28	1	1.0006	0.0496
28×36	1.5	1.4943	0.0731
20×50	2	2.0099	0.1074
20×50	5	4.8894	0.2584
10×100	10	9.6225	0.5927

For the process, solution of the equation in example 2, with parameter $\gamma = 1/2$ and $\gamma = 2$, we get the results in Table 6.

TABLE 6. Mean and standard deviation for the estimate of b in the case of $\gamma = 1/2$ or $\gamma = 2$

		$\gamma = 1/2$			$\gamma = 2$	
b	$k \times m$	Mean($\breve{b}_n$)	SD($\breve{b}_n$)	$k \times m$	Mean($\breve{b}_n$)	SD($\breve{b}_n$)
0.01	100×10	0.0109	0.0025	100×10	0.0141	0.0049
0.1	100×10	0.0999	0.0071	100×10	0.1017	0.0080
0.5	100×10	0.4978	0.0265	100×10	0.5059	0.0267
1	67×15	0.9982	0.0485	100×10	1.0020	0.0521
1.5	50×20	1.4942	0.0742	67×15	1.5057	0.0716
2	50×20	2.0051	0.1033	50×20	2.0065	0.0987
5	28×36	5.0199	0.2478	20×50	5.0353	0.2676
10	15×67	9.8467	0.5368	10×100	10.2960	0.6332

Application 3. In this last application we will study the ergodic process presented in example 3. We study this process to show that our technique works for ergodic processes where the invariant density is an unusuall distribution, in this case, without moments of any order.

The invariant density for the process considered is the Cauchy density, $Cauchy(\alpha, 1)$. For simplicity we take $\sigma = 1$ and deduce the estimator for α using the sub-sampling median.

We obtained,

$$\widehat{\alpha}_n = \frac{1}{m} \sum_{j=1}^{m} Median(X_j, X_{m+j}, ..., X_{(k-1)m+j}). \tag{4.20}$$

We used the Euler scheme in the same conditions as in the previous application. For the same value of α and different combinations of k and m we obtained the results presented in Table 7. We have fixed $\sigma = 1$ and the time lag, $\Delta = 1$.

Table 7. Mean, standard deviation and mean square error for the estimators of α when we use the median estimator in the subsamples

$k \times m$	α	Mean($\widehat{\alpha}_n$)	SD($\widehat{\alpha}_n$)	MSE($\widehat{\alpha}_n$)
10×100	10	10.2209	1.46897	2.20237
25×40	10	10.1789	1.47740	2.21033
40×25	10	10.1784	1.47608	2.20629
100×10	10	10.1720	1.46441	2.16979

In this application we used the median to estimate the centrality parameter of the Cauchy distribution, α. The results we obtained are good and do not depend on the decomposition. In this case we do not present results for other parameter values because the standard deviation and the mean square error were quite similar.

5. Final remarks

This paper points towards the use of the invariant density while converging out inference on ergodic diffusions. The results suggest that the technique provides better estimators than the usual techniques for the considered processes that we were able to compare. It should be emphasized that our estimators not only provide smaller MSE but they are always well defined in contrast with the other presented estimators. For the new processes the results also suggest that the technique is valid when we choose appropriate decomposition of k and m. In fact, an open problem is how to choose the optimal decomposition of the sample. For the Ornstein-Uhlenbeck process this question is answered by proposition 4.2 and the same idea could, eventually, be applied to other processes. For now and for processes where the correlation coefficient equal to zero does not imply independence, we can suggest a simple procedure to obtain a good decomposition. Starting with a default decomposition ($k \approx m$) we obtain initial estimatives for the parameters and with this estimatives we can do simulations to get the optimal decomposition for the same dimension samples and then, we compute estimatives using the original sample and this optimal decomposition. This procedure can be repeated if necessary.

References

[1] R. Durett, *Stochastic Calculus. A Practical Introduction*, CRC Press. New York, 1996.
[2] S. Iacus, *Simulation and Inference for Stochastic Differential Equations with R Examples*, Springer, 2008.
[3] I. Karatzas & S. Shreve, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, 2000.
[4] U. Küchler & M Sørensen, *Exponential Families of Stochastic Processes*, Springer-Verlag, 1997.
[5] Y.A. Kutoyants, *Statistical Inference for Ergodic Diffusion Processes*, Springer-Verlag, 2004.
[6] M. Musiela & M. Rutkowski *Martingale Methods in Financial Modelling*, Springer-Verlag, 2005.
[7] B. Oksendall, *Stochastic Differential Equations*, Springer-Verlag, 1998.
[8] R. Prakasa, *Statistical Inference for Diffusion Type Processes*, Arnold, London and Oxford University press, New York, 1999.
[9] Roger and Williams, *Diffusions, Markov Processes and Martingales*, Cambridge University Press, 2001.
[10] A. Shiryaev, *Essentials of Stochastic Finance: Facts, Models, Theory*, World Scientific, 1999.

Department of Mathematics & CMA of FCT/UNL
E-mail address: lpcr@fct.unl.pt

Department of Mathematics & CMA of FCT/UNL
E-mail address: pjpm@fct.unl.pt

Department of Mathematics & CMA of FCT/UNL
E-mail address: jtpm?@fct.unl.pt