International Journal of Pure and Applied Mathematics
Volume 7 No. 4 2003, 449-484

PARAMETRIC ESTIMATION OF DIFFUSION PROCESSES SAMPLED AT FIRST EXIT TIMES

Jaime A. Londoño
Departamento de Ciencias Básicas
Universidad EAFIT
Carrera 49 # 7 Sur-50, Medellín, COLOMBIA
e-mail: jalondon@sigma.eafit.edu.co

Abstract: This paper introduces a family of recursively defined estimators of the parameters of a diffusion process. We use ideas of stochastic algorithms for the construction of the estimators. Asymptotic consistency of these estimators and asymptotic normality of an appropriate normalization are proved. The results are applied to two examples from the financial literature; viz., Cox-Ingersoll-Ross model and the constant elasticity of variance (CEV) process illustrate the use of the technique proposed herein.

AMS Subject Classification: 62M05, 62P05
Key Words: continuous time Markov processes, discrete time sampling, diffusions, interest rate models, stochastic algorithms

1. Introduction

In this paper we introduce a family of recursively defined estimators of the parameters of a diffusion process. We assume that the process is observed when it reaches some trigger values in a particular order. An extensive literature

Received: April 24, 2003

exists on estimation for a continuous time record of observations of diffusion processes (e.g. Bawasa an Rao [5]). Nelson [50] studies the convergence of stochastic difference equations to stochastic differential equations as the length of discrete time intervals between observations goes to zero. Banon [4] proposes a recursive kernel estimate of an initial density for a stationary Markov process.

The techniques for the use of discretely-observed data are somewhat different from those used for a continuous time record of observations. Maximum likelihood estimation can be applied to discrete data, although most of the current theory requires the discretely sample data to be stationary ergodic Markov chains, see Billingsley [9] or Hall and Heyde [30] for more complete references. Lo [45] derived a functional partial differential equation that characterizes the likelihood function of a discretely sampled Itô process. Likelihood based estimation is usually computationally quite costly because an auxiliary partial differential equation must be solved numerically for each hypothetical parameter value and each observed state. Duffie and Singleton [15] and Gouriéroux et al [29] suggested the use of numerical methods to approximate moments. He [34] proposed the use of binomial approximations. Simulation approaches do not require the Markov state vector to be fully observed. However, it is difficult to determine the magnitude of the approximation error, and in some applications, it might be numerically costly to ensure that the approximation error is small. Hansen and Scheinkman [32] proposed moment conditions suitable for use of generalized method of moment estimators (see Hansen [31]) based on properties of the infinitesimal generators of stationary Markov processes. Aït-Sahalia [1] proposed the use of non-parametric techniques for the estimation of stationary one-dimensional diffusions. Duffie and Glynn [16] introduced a family of generalized method of moments estimators for continuous time Markov processes observed at random time intervals. They assume that the arrival of the data has an intensity that varies with the underlying Markov process or varies with an independent Markov process. An incomplete list of alternative estimation procedures includes Aït-Sahalia [2], Gallant and Tauchen [23], Stanton [54], Bandi and Phillips [3], Chacko and Viceira [11], Singleton [53], Eraker [20] and Jones [37].

Diffusion processes play a fundamental role in stochastic optimal control theory, stochastic thermodynamics, and financial economics. We are particularly interested in models arising from the financial literature. Examples of these are exchange rate models (e.g. see Froot and Obstfeld [22] and Krugman [40]) and models of term structure of interest rates (e.g. see Cox et al [13] and Heath et al [35]).

Our goal is to estimate the parameters of a diffusion process. We obtain results when the state space is one dimensional. We assume that the differential operator of the diffusion process $(X_t, \mathcal{F}_t, \mathbf{P}_x)$ is given by

$$L = \frac{1}{2}\sigma^2(\cdot, \theta^*)\frac{d}{dx^2} + b(\cdot, \theta^*)\frac{d}{dx}\,, \tag{1}$$

where b, σ^2 satisfy some technical conditions sufficient for a diffusion with this differential operator to exist, and $\theta^* \in \mathbb{R}^s$ is a parameter to be estimated. For any twice continuous differentiable function f on $\mathbb{R}$, it is known that

$$Lf(x) = \lim_{U \downarrow \{x\}} \frac{\mathbf{E}_x f(X_{\tau_U}) - f(x)}{\mathbf{E}_x(\tau_U)}\,, \tag{2}$$

where the limit is taken over open sets U containing x and τ_U denotes the first exit time from the open set U. For the precise meaning of equation (2) see Dynkin [19]. Therefore, it is natural to use moment conditions based on the expressions in the numerator and the denominator of equation (2) to construct estimators of the parameters of the diffusion. It turns out that this approach suggests parameterizations that are appropriate for identification of the process from moment conditions of the previous type in a way that is made precise in Section 3. We use ideas from stochastic algorithms for the construction of the estimators. References for the theory of stochastic algorithms are, for instance, Benveniste et al [7], Kushner and Clark [43], Duflo [17], Kushner and Yin [44], Has'Minskii and Nevel'Son [33]. We prove that the sequence of estimators constructed is asymptotically consistent and an appropriate normalization of them is asymptotically normal.

Among the advantages of the technique that we propose are that we do not require the diffusion process to be stationary, to have an invariant probability measure or to satisfy some sort of ergodicity as the techniques in prior works assume. Another nice feature of the estimation that we propose is the computational tractability for any diffusion with continuous differentiable drifts and diffusion coefficients. In fact, we give a closed form for the functions we are required to compute. Finally, there exists an extensive literature developed for the theory of stochastic algorithms, and so it is likely that the ideas used there might be applied to this context. A particularly appealing characteristic of stochastic algorithms in the econometrics of financial time series, as Benveniste et al [6] recall, is its "generally recognized ability to adapt to variations in the underlying systems". The latter could make it useful for the analysis of high frequency data that seems not to be time homogeneous.

The paper is organized as follows: In Section 2 we state some hypotheses that are used subsequently and define the estimators that we propose. In Section 3 we prove the asymptotic consistency of these estimators. In Section 4 we prove that an appropriate normalization of the sequence of estimators defined in Section 2 is asymptotically normal. In Section 5 we show how the theory we develop can be applied to some models of interest rates. Namely, we consider Cox-Ingersoll-Ross model and the constant elasticity of variance (CEV) process.

2. Construction of Estimators

Our goal in this paper is to estimate some parameters of a Markov process using the values of the process that are known at some random times $\tau_1, \tau_2, \tau_3, ...$ and $\nu_2, \nu_3, ...$ that are related to the process through the equations (3) and (5).Genon-Catalot [24],Genon-Catalot and Laredo [25], Genon-Catalot and Laredo [26], Genon-Catalot et al [27], have constructed estimators for the parameters of a diffusion, when only first hitting times are observed. The use of a space discretization rather than time discretization is well known in the probabilistic context; it has been applied in algorithms of path reconstruction. See Kushner and Dupuis [42], Milstein [48] and Milstein and Tretyakov [49].

We assume that we have a parametric set of diffusions indexed by $\mathbb{H} \subset \mathbb{R}^s$, where $\mathbb{H}$ is either a compact set or $\mathbb{H} = \mathbb{R}^s$. We define random variables recursively (see equation (7) and equation (10)) that depend on the data which we observe. Then we prove that, under some technical conditions, the random variables defined in this way are asymptotically consistent for the true value of the parameter when the parameter space is one-dimensional. When the parameter space is multidimensional we obtain convergence to an invariant set of an ODE (Ordinary Differential Equation). From now on we assume that $\Omega = C([0, \infty))$ is the canonical space of continuous S-valued functions where $S = \mathbb{R}$, $S = [0, \infty)$, or $S = (0, \infty)$ with the metrizable topology of uniform convergence on compact sets. We denote by $\mathcal{F}_\infty$ the Borel σ-algebra of Ω. For $0 \leq t < \infty$, X_t is the coordinate mapping process and, $\mathcal{F}_t$ the σ-algebra generated by $X(\cdot)$ on $[0, t]$; namely $\mathcal{F}_t = \{X_s, 0 \leq s \leq t\}$. We define the filtration $\mathcal{F}_{t^+} = \bigcap_{\epsilon > 0} \sigma\left(X_u : \epsilon + t \geq u\right)$, for $t \in [0, \infty)$. For the construction of our estimators we need the following condition:

Condition 1. $(X_t, \mathcal{F}_t, \mathbf{P}_x^\theta)_{\theta \in \mathbb{H}}$ is a parametric set of diffusions with sample space $(\Omega, \mathcal{F}_\infty)$ and differential operators $(L_\theta)_{\theta \in \mathbb{H}}$. For each $\theta \in \mathbb{H} \subset \mathbb{R}^s$, the part of the process $(X_t, \mathcal{F}_t, \mathbf{P}_x^\theta)$ on S is a recurrent strong Markov process. Also,

$$\sup_{x \in (a,b)} \mathbf{E}_x^\theta \, \tau^{(a,b)} < \infty \qquad (a, b) \subset S\,,$$

where $\tau^{(a,b)}$ is the first exit time of the open set (a, b). We assume that the functions $\theta \mapsto \mathbf{E}_x^\theta Z$ and $\theta \mapsto L_\theta f$ are Borel measurable for all $x \in S$, Z a random variable defined on Ω, and $f \in C^2(S)$.

See Dynkin [19, 18] for a definition of recurrence, and of part of a process. In order to guarantee that a given non-negative second order differential operator L defined on $C^2(\mathbb{R})$ is the differential operator of a diffusion process, it is customary to impose:

Condition 2. L is a non-negative second order differential operator with measurable drift coefficient b and *continuous* diffusion coefficient σ^2 that is *uniformly Lipschitz continuous* and satisfies *either* of the following two properties:

1. There exists $c > 0$ such that

$$\sigma^2(x) \geq c \qquad \text{for} \ \ x \in \mathbb{R}\,.$$

2. σ^2 is a twice continuously differentiable function such that the second derivative $d^2\sigma^2/dx^2$ is bounded on $\mathbb{R}$.

If L is as in Condition 2 then there exists a diffusion process $(X_t,\mathcal{F}_t,\mathbf{P}_x)$ whose differential operator is L (see Kunita [41], Corollary 4.2.7.)

Throughout the rest of the paper we shall assume that $(X_t, \mathcal{F}_t, \mathbf{P}_x^\theta)_{\theta\in\mathbb{H}}$ is a parametric set of one-dimensional diffusions with sample space $(\Omega, \mathcal{F}_\infty)$ that satisfies Condition 1 and with differential operators $(L_\theta)_{\theta\in\mathbb{H}}$ that satisfy Condition 2. We assume $\theta^* \in \mathbb{H}$ is a fixed constant and μ is a probability measure supported on S. We denote by $(X_t, \mathcal{F}_t, \mathbf{P})$ the Markov process with initial probability measure μ and probability for paths starting at $x \in S$, $\mathbf{P}_x^{\theta^*}$, namely $\mathbf{P} = \int \mathbf{P}_x^{\theta^*} d\mu(x)$.

We assume that we have a finite set $\mathtt{D} = \{d_1, \ldots, d_s\} \subset S$, where $d_1 < \cdots < d_s$. $\mathtt{D}$ is a set of states, where the process can be observed. We assume that the data arrival process is given by the following sequence of $(\mathcal{F}_{t+})$ stopping times:

$$\begin{aligned} \tau_1^{\mathtt{D}} &= \inf\{t \geq 0 \mid X_t \in \mathtt{D}\}\,, \\ \tau_{n+1}^{\mathtt{D}} &= \inf\{t \geq \tau_n \mid X_t \in \mathtt{D} \setminus \{X_{\tau_n}\}\}\,, \qquad \text{for } n > 1\,. \end{aligned} \tag{3}$$

We suppress $\mathtt{D}$ in what follows. Under the hypothesis of recurrence we observe that (τ_n) is a sequence of finite stopping times.

Let $\{U_d\}_{d\in\mathtt{D}}$ be a finite set of disjoint open connected sets of S such that $U_d \cap \mathtt{D} = \{d\}$ for any $d \in \mathtt{D}$. The boundary points of U_d comprise a set of states, where the process can be observed given that the process has reached the point $d \in \mathtt{D}$. Let $\mathtt{D}_r, \mathtt{D}_l \colon \mathtt{D} \mapsto S \cup \{\infty, -\infty\}$ be the functions satisfying $U_d = (\mathtt{D}_l(d), \mathtt{D}_r(d))$. We define $\eta^f \colon \mathbb{H} \times \mathtt{D} \mapsto \mathbb{R}$ by the formula

$$\eta^f(\theta, x) = \mathbf{E}_x^\theta f(X_{\tau_{\mathtt{D}_r(x)} \wedge \tau_{\mathtt{D}_l(x)}}) \qquad \text{for } x \in \mathtt{D}\,, \tag{4}$$

where $f \colon S \mapsto \mathbb{R}$ is a twice continuous differentiable function. Let

$$\nu_{n+1} = \inf\{t \geq \tau_n \mid X_t \notin U_{X_{\tau_n}}\} \qquad \text{for } n \geq 1\,. \tag{5}$$

We observe that $(\nu_n)_{n \geq 2}$ is a sequence of $(\mathcal{F}_{t+})$ stopping times. We define V^f by the formula:

$$V^f(\theta, x, y) = f(y) - \eta^f(\theta, x)\,. \tag{6}$$

From now on, we shall assume that the set of parameters $\mathbb{H} = \mathbb{R}$ or $\mathbb{H}$ is the constrain set of parameters $\mathbb{H} = \big\{\theta \colon a_i \leq \theta^i \leq b_i\big\}$, $-\infty < a_i < \theta^i < b_i < \infty$ for $1 \leq i \leq s$, where θ^i denotes the i-th component of θ. It is customary in the theory of stochastic algorithms to consider a parameter set that is assumed to be compact, due to the fact that useful parameter values in applications are confined by constrains of physics or economics to some constrain set. The constrain set mentioned above is one of such possibilities. Another alternatives can be considered, see Kushner and Yin [44] or the discussion on Section 3. If $H \colon \mathbb{H} \times \mathtt{D} \to \mathbb{R}^s$ is a measurable map, we define a sequence of estimators by the recursive relation

$$\Theta_{n+1} = \Pi_{\mathbb{H}} \Big[\Theta_n - \gamma_n H(\Theta_n, X_{\tau_n}) V^f(\Theta_n, X_{\tau_n}, X_{\nu_{n+1}})\Big]\,, \tag{7}$$

for $n \geq 1$, where Θ_1 is a bounded random variable taking values in $\mathbb{H}$, $\Pi_{\mathbb{H}}$ is the projection onto $\mathbb{H}$, and (γ_n) is a decreasing sequence of positive numbers with $\gamma_n \downarrow 0$. In particular, Θ_1 can be a constant and $\Pi_{\mathbb{H}} = id$ when $\mathbb{H} = \mathbb{R}$. The meaning of equation (7) is well-known in the theory of stochastic algorithms. A noise corrupted observation $Y_n = H(\Theta_n, X_{\tau_n}) V^f(\Theta_n, X_{\tau_n}, X_{\nu_{n+1}})$ of a vector valued function $\bar{g}(\cdot)$ is taken, whose root $\theta^* \in \mathbb{H}$ we are seeking. Actually, one observes values of the form $Y_n = g(\theta_n, X_{\tau_n}) + \delta M_n$, where δM_n has the property that $\mathbf{E}\left[\delta M_n \mid Y_i, \delta M_i, i < n\right] = 0$. Loosely speaking, Y_n is an "estimator" of $\bar{g}(\cdot)$ in the sense that $\bar{g}(\theta) = \lim_m (1/m) \sum_{i=1}^m g(\theta, X_{\tau_n})$, where $\bar{g}(\cdot)$ is a function based on moment conditions of the type defined by equation (4). The sequence (γ_n) is chosen to provide an implicit average of the iterates.

In a similar way if $g \colon [0, \infty) \mapsto \mathbb{R}$ is a measurable map, we define $\widetilde{\eta}^g \colon \mathbb{H} \times \mathtt{D} \mapsto \mathbb{R}$, $\widetilde{V}^g \colon \mathbb{H} \times \mathtt{D} \times [0, \infty) \mapsto \mathbb{R}$ by the formulas:

$$\widetilde{\eta}^g(\theta, x) = \mathbf{E}_x^\theta\, g(\tau_{\mathtt{D}_r(x)} \wedge \tau_{\mathtt{D}_l(x)})\,, \tag{8}$$

$$\widetilde{V}^g(\theta, x, y) = g(y) - \widetilde{\eta}^g(\theta, x)\,, \tag{9}$$

where we assume that $\{U_d\}_{d\in\mathtt{D}}$ is chosen in such a way that $\widetilde{\eta}^g < \infty$. For instance, under the assumptions of Condition 1 it is enough to assume that the sets $\{U_d\}_{d\in\mathtt{D}}$ have compact closure. As before, if $\widetilde{H}\colon \mathbb{H}\times\mathtt{D} \to \mathbb{R}^s$ is a measurable map, we define a sequence $(\widetilde{\Theta}_n)$ of estimators by the recursive relation

$$\widetilde{\Theta}_{n+1} = \Pi_{\mathbb{H}}\left[\widetilde{\Theta}_n - \gamma_n \widetilde{H}(\widetilde{\Theta}_n, X_{\tau_n})\widetilde{V}^g(\widetilde{\Theta}_n, X_{\tau_n}, \nu_{n+1} - \tau_n)\right], \qquad (10)$$

for $n \geq 1$, where $\widetilde{\Theta}_1$ is a bounded random variable taking values in $\mathbb{H}$. Remarks similar to the ones done for the meaning of equation (7) hold for equation (10).

Instead, of using stochastic algorithms we could estimate the true parameter of the process by trying to minimize the sum of the squares

$$Q_n(\theta) = \sum_{k=1}^{n}\left[f(X_{\nu_{n+1}}) - \mathbf{E}^\theta\left[(f(X_{\nu_{n+1}}) \mid \mathcal{F}_{\tau_n}\right]\right]^2$$
$$= \sum_{k=1}^{n}\left[f(X_{\nu_{n+1}}) - \eta^f(\theta, X_{\tau_n})\right]^2$$

with respect to θ. When the parameter space is unconstrained, the estimates will be taken to be the solution of the system

$$\frac{\partial Q_n(\theta)}{\partial \theta_i} = 0, \qquad \text{for } i = 1, \cdots, s.$$

The "conditional square" approach goes back to Klimko and Nelson [39], among others (see also Hall and Heyde [30]). The methodological reason for which we choose to work with stochastic algorithms instead, it is its recognized ability to adapt to variations of the underlying system, as well as its ability to process data sequentially as they are observed. In the next section we will find conditions that are sufficient for the sequence of random variables defined by equations (7) and (10) to be asymptotically consistent.

3. Consistent Estimation

The next theorem is used to prove asymptotic consistency, in the case of a one-dimensional parameterization, for the sequence of random variables defined in equations (7) and (10), compare with Theorem 7.1 from Kushner and Yin [44].

Theorem 1. (A Robbins-Monro Algorithm) *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space,* $(\mathcal{F}_n)$ *be a filtration of sub-*σ*-algebras of* $\mathcal{F}$, $\mathtt{D} \subset \mathbb{R}$ *be a finite set, and* $(X_n, Y_n, \Theta_n)_{n\in\mathbb{N}}$ *be a sequence of real valued* $(\mathcal{F}_n)$ *adapted random variables, where* X_n *takes values in* $\mathtt{D}$. *Let* Θ_n *be defined by the following recursive relation:*

$$\Theta_{n+1} = \Theta_n - \gamma_n H(\Theta_n, Y_n) V(\Theta_n, Y_n, X_{n+1}) , \tag{11}$$

where $V \colon \mathbb{R} \times \mathtt{D} \times \mathbb{R} \mapsto \mathbb{R}$, $H \colon \mathbb{R} \times \mathtt{D} \mapsto \{1, -1\}$ *are measurable functions,* (γ_n) *is a decreasing sequence of positive numbers and* $\mathbf{E}(\|\Theta_1\|^2) < \infty$. *We assume that the following hypotheses* A_1, A_2, H_1, H_2, *and* H_3 *are satisfied:*

(A_1) *There exist a measurable function* $\overline{V} \colon \mathbb{R} \times \mathtt{D} \mapsto \mathbb{R}$ *such that*

$$\mathbf{E}(V(\Theta_n, Y_n, X_{n+1}) \mid \mathcal{F}_n) = \overline{V}(\Theta_n, Y_n) . \tag{12}$$

(A_2) *There exist a positive measurable function* $S^2 \colon \mathbb{R} \times \mathtt{D} \mapsto [0, \infty)$ *such that*

$$\mathbf{E}(V^2(\Theta_n, Y_n, X_{n+1}) \mid \mathcal{F}_n) = S^2(\Theta_n, Y_n) . \tag{13}$$

(H_1) *There exist* $\theta^* \in \mathbb{R}$ *such that for any* $d \in \mathtt{D}$ *and* $\theta \in \mathbb{R}$

$$(\theta - \theta^*) H(\theta, d) \overline{V}(\theta, d) \geq 0 \tag{14}$$

and there exists an increasing sequence of positive integers $(n_k)_{k\in\mathbb{N}}$ *such that for any* $\varepsilon > 0$

$$\liminf_k \inf_{\varepsilon \leq |\theta - \theta^*|} \mathbf{E}((\theta - \theta^*) H(\theta, Y_{n_k}) \overline{V}(\theta, Y_{n_k})) > 0 . \tag{15}$$

(H_2) *There exist* $K > 0$ *such that*

$$S^2(\theta, d) \leq K(1 + (\theta - \theta^*)^2) \qquad \text{for all } \theta \in \mathbb{R}, \quad d \in \mathtt{D} . \tag{16}$$

(H_3) *The sequence* (γ_n) *of positive numbers satisfies*

$$\sum_{n_k} \gamma_{n_k} = \infty, \qquad \sum_n \gamma_n^2 < \infty . \tag{17}$$

Then the sequence (Θ_n) *converges almost surely to* θ^*.

Typically, a family $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda_1,\lambda_2)})_{(\lambda_1,\lambda_2)\in\mathbb{H}_1\times\mathbb{H}_2}$ of scalar diffusions with differential operators $(L_{(\lambda_1,\lambda_2)})_{(\lambda_1,\lambda_2)\in\mathbb{H}_1\times\mathbb{H}_2}$ is given, where $\mathbb{H}_i \subset \mathbb{R}^{s_i}$, $i = 1, 2$, are compact subsets or $\mathbb{H}_1 \times \mathbb{H}_2 = \mathbb{R}^{s_1} \times \mathbb{R}^{s_2}$, and

$$L_{(\lambda_1,\lambda_2)} = \frac{1}{2}\sigma^2(x, \lambda_1)\frac{d^2}{dx^2} + b(x, \lambda_2)\frac{d}{dx} .$$

It turns out that the sampling structures hinted by equation (7) and equation (10) suggest that it is more natural to assume a parameterization defined by the indexed family of differential operators

$$L_{(\lambda'_1,\lambda'_2)} = \frac{1}{2}\sigma^2(x,\lambda'_1)\frac{d^2}{dx^2} + (b/\sigma^2)(x,\lambda'_2)\sigma^2(x,\lambda'_1)\frac{d}{dx}\,, \tag{18}$$

where $\lambda'_i \in \mathbb{H}'_i$ are compact subsets of $\mathbb{R}^{s_i}$, $i = 1, 2$ or $\mathbb{H}'_1 \times \mathbb{H}'_2 = \mathbb{R}^{s_1} \times \mathbb{R}^{s_2}$, and where σ^2, b/σ^2 are parameterizations of the diffusion and the ratio between the drift and the diffusion respectively (see equations (60) and (62)). It is often the case that the latter parameterization defines an equivalent problem to the former parameterization, at least as estimation is concerned. Indeed, according to Itô and McKean [36] that borrows a phrase of W. Feller, the expression in the numerator of equation (1) defines a road map, i.e. it tells what routes the particle is permitted to travel, and the expression at the bottom of equation (1) defines the speed of the diffusion. Using Feller terminology, λ'_2 identifies the "road map", and λ'_1 identifies the "speed" of the diffusion when the "road map" is known. In this paper we should adopt the latter approach. Corollary 1 below is used to estimate the parameter that identifies the ratio between the drift and the diffusion when the parameter space used to identify this ratio is $\mathbb{R}$. Theorem 2 is used when a compact subset of $\mathbb{R}^{s_1}$ is used as the parameter space that identifies this ratio. Let us observe that neither case requires the parameter(s) that identifies the diffusion to be known. Thus, it is possible to assume that the latter parameter is known, when the estimation of the former parameters of the diffusion is made.

For the next two corollaries, let us assume that the parameter space is one dimensional.

Corollary 1. *Let $\sigma^2 \colon S \times \mathbb{R} \mapsto [0, \infty)$, and $b \colon S \times \mathbb{R} \mapsto \mathbb{R}$ be defined by the formula*

$$L_\lambda = \frac{1}{2}\sigma^2(\cdot,\lambda)\frac{d^2}{dx^2} + b(\cdot,\lambda)\frac{d}{dx}\,, \qquad \text{for } \lambda \in \mathbb{R}\,, \tag{19}$$

where $b(\cdot,\lambda)$, $\sigma^2(\cdot,\lambda)$, $(b/\sigma^2)(\cdot,\lambda) \in C(\overline{S}) \cap C^2(S)$ for any $\lambda \in \mathbb{R}$. We assume that $\partial/\partial\lambda(b/\sigma^2)(x,\lambda)$ exists and is nowhere zero for $(x,\lambda) \in \cup_{d\in\mathtt{D}} U_d \times \mathbb{R}$.

Let $H \colon \mathbb{R} \times \mathtt{D} \to \{-1, 1\}$ be defined by the formula

$$H(\lambda, d) = \mathbf{1}_{(0,\infty)}(\partial/\partial\lambda(b/\sigma^2)(d,\lambda)) - \mathbf{1}_{(-\infty,0)}(\partial/\partial\lambda(b/\sigma^2)(d,\lambda)) \tag{20}$$

If $\lambda^* \in \mathbb{R}$ is a fixed number, then the sequence of random variables defined by equation (7), converges almost surely $\mathbf{P}_x^{\lambda^*}$ to λ^* for any $x \in S$, where η^f and V^f are defined as in equations (4) and (6), and $f =$id is the identity on $\mathbb{R}$.

A few words are needed to review the hypotheses from Corollary 1. If the drift is zero the sampling scheme defined by equation (7) can not be used. In fact, only data obtained using the sampling scheme defined by equation (9) would provide any information (see Corollary 2 below). Within the framework proposed this is indeed natural. If the drift is zero, it is conceivable that only the times between hits of the grids and the end points of the surrounding intervals should provide any information. If $b/\sigma^2(d,\cdot)$, $d \in \mathtt{D}$ are strictly monotone functions around an interval containing the "true" parameter, then it is possible to define a new parameterization that complies with the hypothesis of the previous corollary and allows us to identify the parameter at least from a small interval. Also, Theorem 2 can be used whenever $b/\sigma^2(d,\cdot)$, $d \in \mathtt{D}$ are not strictly monotone.

Corollary 2 below is used to estimate the parameter that identifies the diffusion, when the parameter space to identify this diffusion is $\mathbb{R}$ (see Theorem 2 for estimation of parameters used to identify the diffusion term for a multidimensional setting for the parameter space). It is assumed that the vector of parameter(s) that identifies the ratio between the drift and the diffusion is known. The previous assumption can be made in lieu of Corollary 1 or Theorem 2 in conjunction with the remarks made right after Theorem 1.

Corollary 2. *Let $\sigma^2\colon S \times \mathbb{R} \mapsto [0,\infty)$, and $b\colon S \times \mathbb{R} \mapsto \mathbb{R}$ be defined by the formula*

$$L_\varsigma = \frac{1}{2}\sigma^2(\cdot,\varsigma)\frac{d^2}{dx^2} + b(\cdot,\varsigma)\frac{d}{dx}\,, \qquad \text{for } \varsigma \in \mathbb{R}\,, \tag{21}$$

where $b(\cdot,\varsigma)$, $\sigma^2(\cdot,\varsigma)$, $b/\sigma^2(\cdot,\varsigma) \in C(\overline{S}) \cap C^2(S)$ for any $\varsigma \in \mathbb{R}$. We assume:

1. *There exists a function $s\colon S \mapsto \mathbb{R}$, that does not depend on ς, such that*

$$\frac{b(x,\varsigma)}{\sigma^2(x,\varsigma)} = s(x) \qquad \text{for any } x \in S,\ \varsigma \in \mathbb{R}\,. \tag{22}$$

2. *There exist $\sigma_0\colon \mathbb{R} \mapsto \mathbb{R}^+$, $h\colon S \mapsto \mathbb{R}$ such that*

$$\sigma^2(x,\varsigma) = \sigma_0(\varsigma)h(x) \qquad \text{for any } x \in S,\ \varsigma \in \mathbb{R}\,, \tag{23}$$

   where we assume that σ_0 is a strictly increasing function that is differentiable and

$$\lim_{s\to\infty}\inf_{|\varsigma|\geq s}|\varsigma|\sigma_0(\varsigma) > 0\,. \tag{24}$$

If $\varsigma^* \in \mathbb{R}$ is a fixed number, then the sequence of random variables defined by equation (10) converges almost surely $\mathbf{P}_x^{\varsigma^*}$ to ς^* for any $x \in S$, where $\widetilde{\eta}^g$ and $\widetilde{V}^g$ are defined as in equations (8) and (9) for g the identity on $\mathbb{R}^+$ and $\widetilde{H}\colon \mathbb{R} \times \mathtt{D} \mapsto \{-1, 1\}$ is the constant function equal to 1.

Let us review the hypotheses of Corollary 2. Equation (22) is natural under the assumptions made on the parameterization. See the remarks made after Theorem 1. The factorization of equation (23) often arises in applications. The latter assumption is used to prove monotonicity of the function defined by equation (43). The assumption made on equation (24) can be made without any loss of generality.

In order to illustrate the use of stochastic algorithms for the problem of estimation in the multidimensional case (for the parameter space), we make use of the standard theorem of convergence for truncated stochastic algorithms with correlated noise with step size going to zero. Assume a constrain multidimensional parameter space $\mathbb{H} = \{\theta\colon a_i \le \theta^i \le b_i\}$, $-\infty < a_i < \theta^i < b_i < \infty$ for $1 \le i \le s$. For $\theta \in \mathbb{H}$, define the set $C(\theta)$ as follows. For $\theta \in \mathbb{H}^0$, the interior of $\mathbb{H}$, $C(\theta)$ contains only the zero element; for $\theta \in \partial\mathbb{H}$, the boundary of $\mathbb{H}$, let $C(\theta)$ be the infinite convex cone generated by the outer normals at θ on the faces on which θ lies. Given a continuous $g\colon \mathbb{R}^s \mapsto \mathbb{R}^s$ the *projected* ODE of $\dot{\theta} = g(\theta)$ is defined to be

$$\dot{\theta} = g(\theta) + z, \qquad \theta(t) \in -C(\theta(t))\,,$$

where $z(\cdot)$ is the projection or constrain term, the minimum term needed to keep $\theta(\cdot)$ in $\mathbb{H}$.

Theorem 2. *Let $(Y_n, \Theta_n)_{n\in\mathbb{N}}$ be a sequence of $(\mathcal{F}_{\tau_n})$ adapted measurable maps where $Y_n\colon (\Omega, \mathcal{F}_{\tau_n}) \mapsto (\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and $\Theta_n\colon (\Omega, \mathcal{F}_{\tau_n}) \mapsto (\mathbb{R}^s, \mathcal{B}(\mathbb{R}^s))$. Let K a non singular $s \times s$ matrix. Assume that (Y_n, Θ_n) satisfies the following recursive relation:*

$$\Theta_{n+1} = \Pi_{\mathbb{H}}\left[\Theta_n - \gamma_n K \triangledown \overline{V}(\Theta_n, X_{\tau_n}) V(\Theta_n, X_{\tau_n}, Y_{n+1})\right]\,, \tag{25}$$

where $\Pi_{\mathbb{H}}$ is the projection onto $\mathbb{H}$, $V\colon \mathbb{H} \times \mathtt{D} \times \mathbb{R} \mapsto \mathbb{R}$ is a measurable function, $\overline{V}\colon \mathbb{H} \times \mathtt{D} \mapsto \mathbb{R}$ is a twice continuous differentiable function with differential $\triangledown\overline{V}(\cdot, d)$ for $d \in \mathtt{D}$, K is an invertible matrix and (γ_n) is a decreasing sequence of positive numbers. We assume that

$$\mathbf{E}(V(\Theta_n, X_{\tau_n}, Y_{n+1}) \mid \mathcal{F}_{\tau_n}) = \overline{V}(\Theta_n, X_{\tau_n}) \qquad \text{for } n \ge 1\,. \tag{26}$$

Moreover, it is assumed that the sequence (γ_n), $\gamma_n \downarrow 0$ of positive numbers *satisfies*

$$\sum_n \gamma_n = \infty, \qquad \sum_n \gamma_n^2 < \infty. \tag{27}$$

Then the sequence (Θ_n) *converges almost surely* $\mathbf{P}$ *to an invariant set of the projected ODE*

$$\dot{\theta} = -K\overline{g}(\theta) + z, \qquad \theta(t) \in -C(\theta(t)))\,, \tag{28}$$

for

$$\overline{g}(\theta) = \frac{1}{2}\sum_{d\in \mathtt{D}} p_d \nabla \overline{V}^2(\theta, d)\,,$$

where $p = (p_i)$ *is the left-fixed probability row vector for the Markov chain* $(X_{\tau_n}, \mathcal{F}_{\tau_n})$*. Indeed,* (Θ_n) *converges almost surely to a unique compact and connected component of the set of stationary points of the equation (28). If* θ^* *is an asymptotically stable point of equation (28) and* (Θ_n) *is in some compact set in the domain of attraction of* θ^* *infinitely often with probability* $\geq \rho$*, then* $\Theta_n \to \theta^*$ *with at least probability* ρ*.*

The proof of the above theorem is a straightforward consequence of the Theorem of convergence with probability one for the correlated noise case for stochastic algorithms (see for example Kushner and Yin [44], Theorem 6.1.1). The details of the proof are left to the reader.

More general constrain sets can be consider. For instance, let $q_i(\cdot)$, $i = 1, \cdots, p$ be continuously differentiable real-valued functions on $\mathbb{R}^s$, with gradients $\nabla q_i(\cdot)$, where it is assumed that $\nabla q_i(x) \neq 0$ if $q_i(x) = 0$ and that $\mathbb{H} = \{x \mid q_i(x) \leq 0, i = 1, \cdots, p\}$ is a nonempty, compact connected set. Define $C(x)$ to be the convex cone generated by the set of outward normals $\{\nabla q_i(x) \mid q_i(x) = 0\}$. Suppose that for each x the set $\{\nabla q_i(x) \mid q_i(x) = 0\}$ is either empty or a linear independently set. Then the Theorem 2 remains true with the obvious changes (see Kushner and Yin [44]). Similarly, if $\mathbb{H}$ is a $\mathbb{R}^{s-1}$ dimensional connected compact surface with a continuous differentiable outer normal, and we define $C(x)$, $x \in \mathbb{H}$, to be the linear span of the outer normal at x then Theorem 2 still holds. See also Kushner and Yin [44]. It is worth noting that the former constrain set, as well as the mentioned in Theorem 2, can give rise to new stationary points of the ODE (28), but this is the only type of singular point that can be introduced by the constrains. In many applications when the truncation bounds are large enough, there is only one stationary point θ^* of the ODE (28) that is globally asymptotically stable. Typically, for the kind of application we are heading, $\bar{V}(\theta, d) = \eta(\theta^*, d) - \eta(\theta, d)$, where $\eta\colon \mathbb{H} \times \mathtt{D} \mapsto \mathbb{R}^s$

is a twice continuous differentiable function (on the parameter variable) and $\theta^* \in \mathbb{H}^0$. If

$$\begin{bmatrix} \frac{\partial \eta}{\partial \theta^1}(\theta^*, d_1) & \cdots & \frac{\partial \eta}{\partial \theta^s}(\theta^*, d_1) \\ \vdots & \ddots & \vdots \\ \frac{\partial \eta}{\partial \theta^1}(\theta^*, d_r) & \cdots & \frac{\partial \eta}{\partial \theta^s}(\theta^*, d_r) \end{bmatrix}$$

defines an injection, then θ^* is the unique stationary point of equation (28) in the interior of $\mathbb{H}$, at least for a sufficiently small neighborhood of θ^*.

As a application of Theorem 2, let us assume a family of scalar diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda_1,\lambda_2)})_{(\lambda_1,\lambda_2)\in\mathbb{H}_1\times\mathbb{H}_2}$ with differential operators $(L_{(\lambda_1,\lambda_2)})_{(\lambda_1,\lambda_2)\in\mathbb{H}_1\times\mathbb{H}_2}$ given by equation (18), where $\mathbb{H}_i \subset \mathbb{R}^{s_i}$, $i = 1, 2$ are constrain sets as the ones discussed above, for $i = 1, 2$. It is assumed that there exist a parameter $(\lambda_1^*, \lambda_2^*) \in \mathbb{H}_1 \times \mathbb{H}_2$ such that $(\mathbf{P}_x^{(\lambda_1^*,\lambda_2^*)}) = (\mathbf{P}_x)$. We consider the family of diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda,\lambda_2^*)})_{\lambda\in\mathbb{H}_1}$. Let (Θ_n) be defined by equation (7), where the projection in taken over the set $\mathbb{H}_1$, $\eta^f(\lambda, x) = \mathbf{E}_x^{\lambda,\lambda_2^*} f(X_{\tau_{\mathtt{D}_r(x)}\wedge\tau_{\mathtt{D}_l(x)}})$, $V^f(\lambda, x, y) = f(y) - \eta^f(\lambda, x)$, $\overline{V}^f(\lambda, x) = \eta^f(\lambda_1^*, x) - \eta^f(\lambda, x)$, K is a non-singular matrix and (γ_n) is a sequence as in equation (27). It follows that Theorem 2 applies, and it identifies λ_1^* if this is the unique stationary point of the projected ODE (28). We observe that the computation made to obtain the sequence (Θ_n) does not depend on the value λ_2^* (see Appendix C for the computation of the algorithms). Next, we assume that the parameter λ_1^* is known (the latter can be assumed by the previous remark). We consider the collection of diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda_1^*,\lambda')})_{\lambda'\in\mathbb{H}_2}$. Let $(\tilde{\Theta}_n)$ be the sequence of estimators defined by equation (10), where the projection is taken oven the set $\mathbb{H}_2$, $\eta^g(\lambda', x) = \mathbf{E}^{\lambda_1^*,\lambda'} g(\tau_{\mathtt{D}_r(x)} \wedge \tau_{\mathtt{D}_l(x)})$, $\tilde{V}^g(\theta, x, y) = g(y) - \tilde{\eta}^g(\theta, x)$, $\overline{V}^g(\lambda', x) = \tilde{\eta}^g(\lambda_2^*, x) - \tilde{\eta}^g(\lambda, x)$, K a non-singular matrix (not necessarily identical to the one used to compute (Θ_n)), then Theorem 2 applies, and it identifies λ_2^* if this is the unique stationary point of the projected ODE (28).

It is worth noting that even if only data associated with the sampling scheme related with equation (7) is available then at least, identification of the the ratio between the drift and the diffusion can be made. Also, when the dimension of the parameter space that identifies either the diffusion or the ratio between the drift and the diffusion are one-dimensional, Corollary 1 and Corollary 2 can be called for the estimation with the advantage that complete identification of the parameter is easier.

4. Asymptotic Normality

In this section we propose a version of the central limit theorem for the class of estimators of Theorem 1.

For any stopping time τ we denote as θ_τ the measurable map defined as $\theta_\tau(\omega)(\cdot) = \theta(\tau(\omega) + \cdot)$. We observe that $\theta_{\tau_n} = \theta_{\tau_2}^{n-1}$ and $X_{\tau_n} = X_{\tau_1} \circ \theta_{\tau_n} = X_{\tau_1} \circ \theta_{\tau_2}^{n-1}$ for $n \geq 2$.

For the following theorem we assume that $(X_t,\mathcal{F}_t,\mathbf{P}_x^\theta)_{\theta\in\mathbb{R}}$ is a parametric family of recurrent strong Markov processes .

Theorem 3. *Let $Y\colon(\Omega,\mathcal{F}_{\tau_2}) \to (\mathbb{R},\mathcal{B}(\mathbb{R}))$ be a measurable map that is bounded below. Moreover, assume that $Y \in \bigcap_{d\in\mathtt{D}} L^2(\mathbf{P}_d)$. Let Y_n be defined as $Y_n = Y \circ \theta_{\tau_{n-1}} = Y \circ \theta_{\tau_2}^{n-2}$, for $n \geq 3$ and $Y_2 = Y$. Let $\eta\colon \mathbb{R}\times\mathtt{D} \to \mathbb{R}$ be the function defined as $\eta(\theta, d) = \mathbf{E}_d^\theta(Y)$. In addition assume that Hypotheses N_1 and N_2 are satisfied:*

(N$_1$) *For any $d \in \mathtt{D}$ $\eta(\cdot, d)$ is a strictly monotone, twice continuous differentiable function with non-vanishing derivative.*

(N$_2$) *There exist $L, L' > 0$ such that for any $\theta \in \mathbb{R}$, $d \in \mathtt{D}$*

$$\mid \eta(\theta, d) - \eta(\theta^*, d) \mid\leq L \mid \theta - \theta^* \mid + L' . \tag{29}$$

Define $V\colon \mathbb{R}\times\mathtt{D}\times\mathbb{R} \to \mathbb{R}$ by the formula $V(\theta, d, y) = y - \eta(\theta, d)$. Assume that Θ_n^N is a $(\mathcal{F}_{\tau_n})$ adapted sequence of random variables that satisfies the recursive relation:

$$\Theta_{n+1}^N = \Theta_n^N - \frac{1}{n}\frac{V(\Theta_n^N, X_{\tau_n}, Y_{n+1})}{\alpha(X_{\tau_n})}, \tag{30}$$

where $\alpha(d) = -(\partial\eta/\partial\theta)(\theta^*, d)$ for $d \in \mathtt{D}$ and $\mathbf{E}((\Theta_1)^2) < \infty$.

Then $n^{1/2}(\Theta_n^N - \theta^*)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 = \sum_{d\in\mathtt{D}} p_d Var_d(Y)/\alpha^2(d) = \sum_{d\in\mathtt{D}} p_d \mathbf{E}_d(Y - \eta(\theta^*, d))^2/\alpha^2(d)$, where $p = (p_i)$ is the left-fixed probability row vector of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ as in Lemma 1.

Corollary 3. *Let μ be a probability measure $\mathbb{R}$ supported on S, $\lambda^* \in \mathbb{R}$ be a fixed constant, and let $(\Omega, \mathbf{P}, \mathcal{F}_\infty)$ be the probability space, where $\mathbf{P} = \int \mathbf{P}_x^{\lambda^*} d\mu(x)$. Let $b, \sigma^2, b/\sigma^2$ and (L_λ) be as in Corollary 1. Let $\eta\colon \mathbb{R}\times\mathtt{D} \to \mathbb{R}$ be defined as $\eta(\lambda, d) = \mathbf{E}_d^\lambda(X_{\nu_2})$. We define $V\colon \mathbb{R}\times\mathtt{D}\times\mathbb{R} \to \mathbb{R}$ by the formula $V(\lambda, d, y) = y - \eta(\lambda, d)$. Assume that (Θ_n^N) is a $(\mathcal{F}_{\tau_n})$ adapted sequence of random variables, which satisfies the recursive relation*

$$\Theta_{n+1}^N = \Theta_n^N - \frac{1}{n}\frac{V(\Theta_n^N, X_{\tau_n}, X_{\nu_{n+1}})}{\alpha(X_{\tau_n})}, \tag{31}$$

where $\alpha(d) = -(\partial\eta/\partial\lambda)(\lambda^*, d)$ for $d \in \mathtt{D}$ and Θ_1^N is a bounded random variable. Then $n^{1/2}(\Theta_n^N - \lambda^*)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 = \sum_{d\in\mathtt{D}} p_d \mathbf{E}_d^{\lambda^*}(X_{\nu_2} - \eta(\lambda^*, d))^2/\alpha^2(d)$, where $p = (p_i)$ is the left-fixed probability row vector of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ as in Lemma 1.

Corollary 4. *Let μ be a probability measure on $\mathbb{R}$ supported on S, $\varsigma^* \in \mathbb{R}$ be a fixed constant, and let $(\Omega, \mathbf{P}, \mathcal{F}_\infty)$ be the probability space, where $\mathbf{P} = \int \mathbf{P}_x^{\varsigma^*} d\mu(x)$. We assume that b, σ^2, (L_ς), s, σ_0, and h satisfy the hypothesis of Corollary 2. Let $\tilde\eta \colon \mathbb{R} \times \mathtt{D} \to \mathbb{R}$ be defined as $\tilde\eta(\varsigma, d) = \mathbf{E}_d^{\varsigma}(\nu_2)$. We define $\tilde V \colon \mathbb{R}\times\mathtt{D}\times\mathbb{R} \to \mathbb{R}$ by the formula $\tilde V(\varsigma, d, y) = y - \tilde\eta(\varsigma, d)$. Assume that $(\tilde\Theta_n^N)$ is a $(\mathcal{F}_n)$ adapted sequence of random variables that satisfies the recursive relation*

$$\tilde\Theta_{n+1}^N = \tilde\Theta_n^N - \frac{1}{n}\frac{\tilde V(\tilde\Theta_n^N, X_{\tau_n}, \nu_{n+1} - \tau_n)}{\tilde\alpha(X_{\tau_n})}\,, \tag{32}$$

where $\tilde\alpha(d) = -(\partial\tilde\eta/\partial\varsigma)(\varsigma^*, d)$ for $d \in \mathtt{D}$ and $\tilde\Theta_1^N$ is a bounded random variable. Then $n^{1/2}(\tilde\Theta_n^N - \varsigma^*)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 = \sum_{d\in\mathtt{D}} p_d \mathbf{E}_d^{\varsigma^*}(\nu_2 - \tilde\eta(\varsigma^*, d))^2/\tilde\alpha^2(d)$ where $p = (p_i)$ is the left-fixed probability row vector of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ as in Lemma 1.

In order to illustrate the use of stochastic algorithms, for the problem of asymptotic normality in the multidimensional case (for the parameter space), we use a standard theorem for the rate of convergence for stochastic algorithms with correlated noise and decreasing step size. We assume a constrain multi-dimensional parameter space $\mathbb{H} = \{\theta \colon a_i \le \theta^i \le b_i\}$, $-\infty < a_i < \theta^i < b_i < \infty$ for $1 \le i \le s$. We assume that $(X_t, \mathcal{F}_t, \mathbf{P}_x^\theta)_{\theta\in\mathbb{H}}$ is a parametric family of recurrent strong Markov processes with sample space $(\Omega, \mathcal{F}_\infty)$. Let $\theta^* \in \mathbb{H}$ be an interior point, and μ is a probability measure supported on S. We denote as $(X_t, \mathcal{F}_t, \mathbf{P})$ the Markov process with parameter θ^* and initial probability measure μ. Let Y, Y_n $n \ge 2$ be defined as in Theorem 3, and define $\eta \colon \mathbb{H} \times \mathtt{D} \to \mathbb{R}$ as $\eta(\theta, d) = \mathbf{E}_d^\theta(Y)$. We assume a recursive sequence of estimators $(\mathit{\Theta}_n)$ defined as,

$$\mathit{\Theta}_{n+1} = \Pi_{\mathbb{H}}\left[\mathit{\Theta}_n + \frac{1}{n}K\nabla\eta(\mathit{\Theta}_n, X_{\tau_n})(Y_{n+1} - \eta(\mathit{\Theta}_n, X_{\tau_n}))\right]. \tag{33}$$

Let $D(-\infty,\infty)$ (resp., $D[0,\infty)$) denote the space of real-valued functions on the interval $(-\infty,\infty)$ (resp., on $[0,\infty)$) that are right continuous and have left hand limits, endowed with the Skorohod topology, and $D^s(-\infty,\infty)$ its s-fold product. Full descriptions and treatments of the Skorohod topology are given in Billingsley [10] and Ethier and Kurtz [21]. Define $U_n = \sqrt{n}(\mathit{\Theta}_n - \theta^*)$, and let $U^n(\cdot)$ denote the piecewise constant right continuous interpolation (with interpolation intervals $\{1/n\}$) of the sequence $\{U_i, i \geq n\}$ on $[0,\infty)$. Namely, if we define $t_0 = 0$ and $t_n = \sum_{i=1}^{n} 1/i$, we make

$$U^n(t) = U_m \text{ for } t \in [t_{m-1} - t_{n-1}, t_m - t_{n-1})\,, \text{ and } m \geq n \geq 1\,.$$

For $t \geq 0$, let $m(t)$ denote the unique value of n such that $t \in [t_{n-1}, t_n)$, and for $t < 0$ set $m(t) = 1$. Define the continuous time interpolation $W^n(\cdot)$ on $(-\infty, \infty)$, for $n \geq 1$, by

$$W^n(t) = \begin{cases} \sum_{i=n}^{m(t_n+t)-1} \frac{1}{\sqrt{i}} \left(K \triangledown \eta(\theta^*, X_{\tau_i})(Y_{i+1} - \eta(\theta^*, X_{\tau_i}))\right) & \text{for t } \geq 0\,, \\ -\sum_{m(t_n+t)}^{n} \frac{1}{\sqrt{i}} \left(K \triangledown \eta(\theta^*, X_{\tau_i})(Y_{i+1} - \eta(\theta^*, X_{\tau_i}))\right) & \text{for } t < 0\,. \end{cases} \tag{34}$$

Theorem 4. *Let Y, and Y_n be defined as in Theorem 3. Assume the algorithm given by equation* (33)*, where K is a nonsingular symmetric positive definite matrix. Let θ^* be an isolated stable point of the ODE* (28) *in the interior of $\mathbb{H}$, and assume that (Θ_n) converges almost surely to the process with constant value θ^*. Assume that $\eta(\cdot, d)$, defined as above for $d \in \mathtt{D}$ are twice continuous differentiable functions. Assume that the Hessian matrix*

$$A = D^2_{\theta^*}\Big(\frac{1}{2}\sum_{d\in\mathtt{D}} p_d(\eta(\theta, d) - \eta(\theta^*, d))^2\Big)$$

is positive definite. Moreover, assume that the eigenvalues of the matrix KA are greater that $1/2$ *(in particular the matrix $(-KA+I/2)$ is negative definite). Then, the sequence $(U^n(\cdot), W^n(\cdot))$ converges weakly in $D^r\,[0,\infty) \times D^r\,(-\infty,\infty)$ to a limit process $(U(\cdot), W(\cdot))$, where $W(\cdot)$ is a Wiener process with covariance matrix*

$$\Sigma = \sum_{d\in\mathtt{D}} p_d \mathrm{Var}_d(Y)(K\triangledown(\theta^*, d))(K\triangledown(\theta^*, d))'$$

and $U(\cdot)$ a stationary process with

$$\begin{aligned} U(t) &= U(0) + \textstyle\int_0^t (-KA + I/2)U(s)\,ds + W(t) \\ &= \textstyle\int_{-\infty}^{t} \exp\left((-KA + I/2)(t-s)\right) dW(s) \end{aligned}$$

The proof of the previous theorem follows in a straightforward manner from the rate of convergence theorem for stochastic algorithms with exogenous noise and decreasing step size (see for instance [44], Theorem 10.2.2). Details are left to the reader. A few words are needed to review the hypothesis of the previous theorem. Theorem 2 above, gives sufficient conditions for the almost surely convergence of the sequence (Θ_n). If θ^* is a unique stationary point of the ODE (28) in the interior of $\mathbb{H}$, and K, A are positive definite symmetric matrices then θ^* is a stable point of the ODE. The latter follows from the basic theory of Dynamical systems (see for instance Perko [51]). It is worth noting that Theorem 4 requires the existence of a unique stationary point of the ODE in the interior of $\mathbb{H}$. In the setting of Theorem 4, a sufficient condition for the uniqueness in the interior of $\mathbb{H}$ is discussed after Theorem 2. For any matrix A, as above, it is clearly possible to find a symmetric, positive definite matrix "large enough" so that the eigenvalues of the matrix are greater that 1/2.

Among the choices for K in equation (33) the asymptotic optimal covariance is achieved by $K = A^{-1}$. For this, the limit $U(\cdot)$ satisfies

$$dU = (-KA + I/2))\, U\, dt + K\Sigma^{1/2} dW_0 \,,$$

where W_0 is the standard Winner process. The stationary covariance is

$$\int_0^\infty e^{(-KA+I/2)t} K\Sigma K' e^{(-A'K'+I/2)t}\, dt$$

the trace of this matrix is minimized by choosing $K = A^{-1}$, which yields the asymptotically covariance $A^{-1}\Sigma(A')^{-1}$ (see Kushner and Yin [44] for a deeper discussion on the latter). In order to determine a choice of K that is optimal for the class of estimators proposed by equation (33) it is necessary to have a consistent estimator for the parameter θ^*. This can be accomplished by initially employing a not necessarily optimal estimator of the type mentioned above. As we mentioned earlier, in Section 3, it is possible to make use of Theorem 4 for the estimation of the parameters related with the diffusion or the parameters related with the ratio between the drift and the diffusion.

5. Examples

In this section we show as an illustration the estimation for two examples from the financial literature. We choose the estimation for the one-dimensional parameter space. Details for the multidimensional parameter space are left to the reader to fill out.

5.1. CEV

In this example we consider the estimation of the parameters for the constant elasticity of variance (CEV) process introduced by Cox [12] and by Cox and Ross [14]. The application of this process to interest rates is discussed in Marsh and Rosenfeld [46].

Let us assume that $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda,\varsigma)})_{(\lambda,\varsigma)\in\mathbb{R}^2}$ is a parametric set of diffusions on $\mathbb{R}$ with sample space $(\Omega, \mathcal{F}_\infty)$. We assume that the differential operator of the diffusion $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda,\varsigma)})$ is given by the formula

$$L_{\lambda,\varsigma} = \frac{1}{2}\sigma^2(\varsigma)x^{2\gamma}\frac{d^2}{dx^2} + (\mu/\sigma^2)(\lambda)\sigma^2(\varsigma)\frac{d}{dx} \qquad \text{for } \lambda, \varsigma \in \mathbb{R}\,,$$

where $\mu/\sigma^2 \colon \mathbb{R} \mapsto \mathbb{R}^+$ is a differentiable function such that $d(\mu/\sigma^2)/d\lambda > 0$ and $\sigma^2 \colon \mathbb{R} \mapsto \mathbb{R}^+$ is a continuous differentiable function with $d(\sigma^2)/d\varsigma > 0$ that satisfies equation (24), and $\gamma > 1$ is a fixed known constant. Let (λ^*, ς^*) be the true parameter of the process. Let $\mathtt{D} = \{d_1, \dots, d_s\} \subset \mathbb{R}$ be a finite set of real numbers where $d_1 < \cdots < d_s$. First, we consider the collection of diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda,\varsigma^*)})_{\lambda\in\mathbb{R}}$. Let $f \colon \mathbb{R} \to \mathbb{R}$ be the identity function, and let η^f, V^f be defined as in equations (4) and (6) respectively. We suppress f in what follows. It follows that

$$\eta(\lambda, x) = \{\mathtt{D}_r(x) - \mathtt{D}_l(x)\}\frac{\int_{\mathtt{D}_l(x)}^{x} \exp(\frac{\mu/\sigma^2(\lambda)}{\gamma-1}y^{2-2\gamma})\ dy}{\int_{\mathtt{D}_l(x)}^{\mathtt{D}_r(x)} \exp(\frac{\mu/\sigma^2(\lambda)}{\gamma-1}y^{2-2\gamma})\ dy} + \mathtt{D}_l(x)\,,$$

for $x \in \mathtt{D}$. Let (Θ_n) and (Θ_n^N) be the $(\mathcal{F}_{\tau_n})$ adapted sequences of random variables defined as in equations (7) and (31), respectively, where $H \colon \mathbb{R} \times \mathtt{D} \to \{1, -1\}$ is the constant function taking value -1, $\alpha(d) = -(\partial\eta/\partial\lambda)(\lambda^*, d)$ for $d \in \mathtt{D}$ and Θ_1 is a finite $\mathcal{F}_1$ measurable random variable. We observe that the computation of the estimators Θ_n does not depend on the value of ς^*.

It follows by Corollary 1 that the sequence of estimators (Θ_n) converges almost surely $\mathbf{P}_x^{(\lambda^*,\varsigma^*)}$ to λ^*. Let μ be a probability measure on S and let $(\Omega, \mathbf{P}, \mathcal{F}_\infty)$ be the sample space, where $\mathbf{P} = \int \mathbf{P}_x^{(\lambda^*,\varsigma^*)} d\mu(x)$. As a consequence of Corollary 3, $n^{1/2}(\Theta_n^N - \lambda^*)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 = \sum_{d\in\mathtt{D}} p_d \mathbf{E}_d^{(\lambda^*,\varsigma^*)}(X_{\nu_2} - \eta(\lambda^*, d))^2/\alpha^2(d)$, where $p = (p_i)$ is the left-fixed probability row vector of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ as in Lemma 1.

Next, we assume that the parameter λ^* is known (the later can be assumed by the previous remark.) We consider the collection of diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda^*,\varsigma)})_{\varsigma\in\mathbb{R}}$.

We define $\tilde{\eta}^g$ and $\tilde{V}^g$ by equations (8) and (9) respectively where g is the identity function. See Lemma 7 of Appendix B for a computation of these functions. We suppress g in what follows. Let $(\tilde{\Theta}_n)$, $(\tilde{\Theta}_n^N)$ be the $(\mathcal{F}_{\tau_n})$ adapted sequence of random variables as in equations (10) and (32) respectively, where $\tilde{\Theta}_1$ is a bounded $\mathcal{F}_1$ random variable, $\tilde{H}\colon \mathbb{R}\times\mathtt{D} \mapsto \{-1,1\}$ is the constant function equal to 1 and $\tilde{\alpha}(\cdot) = -(\partial\tilde{\eta}/\partial\varsigma)(\varsigma^*,\cdot)$. It follows by Corollary 2 that for any $x \in \mathbb{R}$ the sequence $(\tilde{\Theta}_n)$ converges almost surely $\mathbf{P}_x^{(\lambda^*,\varsigma^*)}$ to ς^*. It follows that the sequence of estimators $(\tilde{\Theta}_n)$ converges almost surely $\mathbf{P}_x^{(\lambda^*,\varsigma^*)}$ to ς^*. Next, it follows by Corollary 4, that $n^{1/2}(\tilde{\Theta}_n^N - \varsigma^*)$ is asymptotically normally distributed with mean zero and variance $\sigma^2 = \sum_{d\in\mathtt{D}} p_d \mathbf{E}_d^{(\lambda^*,\varsigma^*)}(\nu_2 - \tilde{\eta}(\varsigma^*,d))^2/\tilde{\alpha}^2(d)$.

5.2. Cox-Ingersoll-Ross

In this example we consider the estimation of some parameters for the model of term structure of interest rates of Cox-Ingersoll-Ross (see Cox et al [13]). We consider the problem of the estimation of the quotient between the "speed of adjustment" and the "volatility of the process" (see Cox et al [13] for explanation of this terminology). Let us assume that $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda,\varsigma)})_{(\lambda,\varsigma)\in\mathbb{R}^2}$ is a parametric set of diffusions on $\mathbb{R}$ with sample space $(\Omega, \mathcal{F}_\infty)$. Let (λ^*, ς^*) be the "true" parameter of the process. We assume that the differential operator of the diffusion $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda,\varsigma)})$ is given by the formula

$$L_{\lambda,\varsigma} = \frac{1}{2}\sigma^2(\varsigma)\, x \frac{d^2}{dx^2} + (\mu/\sigma^2)(\lambda)\sigma^2(\varsigma)(\alpha - x)\frac{d}{dx} \qquad \text{for } \lambda, \varsigma \in \mathbb{R}\,,$$

where $\alpha \in \mathbb{R}^+$ is a given known constant, $\mu/\sigma^2\colon \mathbb{R} \mapsto \mathbb{R}^+$ is a differentiable function with $d(\mu/\sigma^2)/d\lambda > 0$ and $\sigma^2\colon \mathbb{R} \mapsto \mathbb{R}^+$ is a continuous differentiable function with $d(\sigma^2)/d\varsigma > 0$ that satisfies equation (24). Let $\mathtt{D} = \{d_1, \ldots, d_s\} \subset (0,\alpha) \cup (\alpha,\infty)$ be set of positive real numbers such that $d_1 < \cdots < d_s$. First, we consider the collection of diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{\lambda,\varsigma^*})_{\lambda\in\mathbb{R}}$. Let $f\colon \mathbb{R} \to \mathbb{R}$ be the identity and let η^f, V^f be defined as in equations (4) and (6) respectively. We suppress f in what follows. Moreover, we assume that $U_d \subset [0,\alpha) \cup (\alpha,\infty)$ for $d \in \mathtt{D}$. Although a diffusion with differential operator as above is not a regular Markov process on $\mathbb{R}$, we observe that if $1/2\alpha > (\mu/\sigma^2)(\lambda)$ $(1/2\alpha \le (\mu/\sigma^2)(\lambda))$ then the part of the process on $[0,\infty)$ (on $(0,\infty)$) (see [19], Volume I for a definition of the part of a process on a subset of the state space) is a regular diffusion on $[0,\infty)$ (on $(0,\infty)$) in the sense of definition 15.1 of Dynkin [19, Volume II, p. 121]. For a discussion of this see, for example Cox et al [13]. Either case follows from the analysis of the boundary classification criteria, see for instance Gihman and A.V. Skorohod [28]. In either case Condition 1 is satisfied if $a, b \in (0,\infty)$. It follows that

$$\eta^f(\lambda, x) = \{\mathtt{D}_r(x) - \mathtt{D}_l(x)\} \times \frac{\int_{\mathtt{D}_l(x)}^{x} y^{-2\alpha(\mu/\sigma^2)(\lambda)} \exp(2(\mu/\sigma^2)(\lambda)y)\,dy}{\int_{\mathtt{D}_l(x)}^{\mathtt{D}_r(x)} y^{-2\alpha(\mu/\sigma^2)(\lambda)} \exp(2(\mu/\sigma^2)(\lambda)y)\,dy} + \mathtt{D}_l(x)\,,$$

for $x \in \mathtt{D}$. Let Θ_1 be a finite $\mathcal{F}_1$ random variable, and let (Θ_n) be the $(\mathcal{F}_{\tau_n})$ adapted sequence of random variables defined as in equation (7), where $H : \mathbb{R} \times \mathtt{D} \to \{1, -1\}$ is the function defined by the formula $H(\lambda, d) = \mathbf{1}_{(-\infty,\alpha)}(d) - \mathbf{1}_{(\alpha,\infty)}(d)$. If $\lambda^* \in \mathbb{R}$ is a fixed number it follows by Corollary 1 that the sequence (Θ_n) of random variables converges to λ^* almost surely $\mathbf{P}_x^{(\lambda^*,\varsigma^*)}$ for any $x \in (0,\infty)$. Next, we assume that the parameter λ^* is known. We consider the collection of diffusions $(X_t, \mathcal{F}_t, \mathbf{P}_x^{(\lambda^*,\varsigma)})_{\varsigma\in\mathbb{R}}$. We define $\tilde{\eta}^g$ and $\tilde{V}^g$ by equations (8) and (9) respectively, where g is the identity function, $\mathtt{D} \subset (0,\infty)$ and $U_d \subset (0,\infty)$. Let $(\tilde{\Theta}_n)$ be a sequence of random variables as in equation (10), where $\tilde{\Theta}_1$ is a bounded random variable and $\tilde{H} : \mathbb{R} \times \mathtt{D} \mapsto \{-1, 1\}$ is the constant function equal to 1. It follows by Corollary 2 that for any $\varsigma^* \in \mathbb{R}$, the sequence $(\tilde{\Theta}_n)$ converges almost surely $\mathbf{P}_x^{(\lambda^*,\varsigma^*)}$ to ς^* for any $x \in (0,\infty)$.

Similar considerations can be made if we want to estimate the "central location" or "long term value" of the process and the diffusion coefficient (see Cox et al [13] for a explanation of this terminology).

Last, we notice that as a consequence of Corollary 3 and Corollary 4 asymptotic normality of the appropriate normalization of the estimators constructed for the Cox-Ingersoll-Ross model can be obtained. The details are left to the reader.

6. Conclusion

The thrust of this paper has been to introduce the ideas of stochastic algorithms to the problem of the estimation of parameters of a continuous diffusion process using observed discrete data. The later could be potentially useful for the study of non time-homogeneous diffusion processes. Besides, we have proposed sampling schemes that depends on space discretization rather than time discretization. These sampling schemes are closer to the Markov character of diffusion processes. We, also, propose a new parameterization of diffusions that we believe is closer in spirit to the initial attempts made in probability to describe a diffusion by its "road map" and "speed". The main results given here (construction of sequences of estimators, asymptotic consistency of such sequences, and asymptotic normality of such sequences), as well as the two examples taken from *Mathematical Finance* dealt with families of diffusion processes that have a one-dimensional state space and a multidimensional parameter space.

Future questions will center on the generalization of the current techniques for use in the case of a multi-dimensional state space and the development for the current setting of stochastic algorithms appropriate to the description of non time-homogeneous diffusions. A particularly interesting question is to find sufficient conditions on multi-dimensional parameterizations of diffusion operators that guarantee identification of the corresponding process from moment conditions of the type presented in this paper. Another direction of research can be centered on the the effective computation of the stochastic algorithms presented here and its comparison with other techniques.

Acknowledgement

I am very grateful to my advisor Professor Neil E. Gretsky for his patience and constructive suggestions. I, also, want to thank Professor Darrel Duffie for his suggestion about my area of research. All remaining errors are mine.

References

[1] Y. Aït-Sahalia, Nonparametric pricing of interest rate derivative securities, *Econometrica*, **64**, No. 3 (1996), 527–560.

[2] Y. Aït-Sahalia, Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach, *Econometrica*, **70**, No. 1 (2002), 223–262.

[3] F. Bandi, P.C. Phillips, *Econometric estimation of diffusion models*, Working Paper (1999).

[4] G. Banon, Nonparametric identification for diffusion processes, *SIAM Journal of Control and Optimization*, **16**, No. 3 (1978), 380–395.

[5] I.V. Basawa, B.L.S. Prakasa Rao, *Statistical Inference for Stochastic Processes*, Probability and Mathematical Statistics, Academic Press, London-New York-Toronto-Sydney-San Francisco (1980).

[6] A. Benveniste, M. Métivier, P. Priouret, Adaptive algorithms and stochastic approximations, *Applications of Mathematics*, **22**, Springer-Verlag, Berlin-Heidelberg-New York (1987); Title of the original French edition: Algorithmes adaptatifs et approximations stochastiques.

[7] A. Benveniste, M. Métivier, P. Priouret, Adaptive algorithms and stochastic approximations, *Applications of Mathematics*, **22**, Springer-Verlag, Berlin (1990), Translated from the French by Stephen S. Wilson.

[8] P. Billingsley, The Lindeberg-Lévy theorem for martingales, *Transactions of the American Mathematical Society*, **12** (1961), 788–792.

[9] P. Billingsley, *Statistical Inference for Markov Processes*, Statistical Research Monographs, **2**, The University of Chicago Press, Chicago (1961).

[10] P. Billingsley, *Convergence of Probability Measures*, Second Edition, Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley and Sons Inc., New York (1999).

[11] G. Chacko, L.M. Viceira, *Spectral gmm Estimation of Continuous-Time Processes*, Working Paper (1999).

[12] J.C. Cox, *Notes on Option Pricing I: Constant Elasticity of Variance Diffusions*, Working paper (1975).

[13] J.C. Cox, J.E. Ingersoll, Jr., S.A. Ross, A theory of the term structure of interest rates, *Econometrica*, **53**, No. 2 (1985), 385–407.

[14] J.C. Cox, S.A. Ross, The valuation of options for alternative stochastic processes, *Journal of Financial Economics*, **3** (1976), 145–166.

[15] D. Duffie, K.J. Singleton, Simulated moment estimation of Markov models of asset prices, *Econometrica*, **61**, No. 4 (1993), 929–952.

[16] D. Duffie, P. Glynn, Estimation of continuous-time Markov processes sampled at random time intervals, Preliminary draft (1996).

[17] M. Duflo, *Random Iterative Models*, Applications of Mathematics, **34**, Springer-Verlag, Berlin (1997); Translated from the 1990 French original by Stephen S. Wilson and revised by the author.

[18] E.D. Dynkin, *Foundations of the Theory of Markov Processes*, Pergamon Press, Oxford, London-New York-Paris (1960), English translation.

[19] E.D. Dynkin, *Markov Processes*, Volumes: **I**, **II**, Springer, Verlag, Berlin-Gottingen-Heidelberg (1965).

[20] B. Eraker, *Mcmc Analysis of Diffusion Models with Applications to Finance*, Working Paper (1997).

[21] S.N. Ethier, T.G. Kurtz, *Markov processes. Characterization and Convergence*, Wiley series in probability and mathematical statistics, John Wiley & Sons, Inc., New York-Chichester-Brisbane-Toronto-Singapore (1985).

[22] K. Froot, M. Obstfeld, Exchange rate dynamics under stochastic regime shifts: a unified approach, *Journal of International Economics*, **31**, No. 3-4 (1991), 203–229.

[23] A.R. Gallant, G. Tauchen, Which moments to match?, *Econometric Theory*, **12**, No. 4 (1996), 657–681.

[24] V. Genon-Catalot, Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21**, No. 1 (1990), 99–116.

[25] V. Genon-Catalot, C. Laredo, Limit theorems for the first hitting times process of a diffusion and statistical applications, *Scand. J. Statist.*, **14**, No. 2 (1987), 143–160.

[26] V. Genon-Catalot, C. Laredo, An asymptotic sufficiency property of observations related to the first hitting times of a diffusion, *Statist. Probab. Lett.*, **9**, No. 3 (1990), 253–258.

[27] V. Genon-Catalot, C. Laredo, M. Nussbaum, Asymptotic equivalence of estimating a Poisson intensity and a positive diffusion drift, *Ann. Statist.*, **30**, No. 3 (2002), 731–753; Dedicated to the memory of Lucien Le Cam.

[28] I.I. Gihman, A.V. Skorohod, *Stochastic Differential Equations*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Volume **72**, Springer-Verlag, New York-Heidelberg-Berlin (1972).

[29] C. Gouriéroux, A. Monfort, E. Renault, Indirect inference, *Journal of Applied Econometrics*, **8** (1993), S85–S118.

[30] P. Hall, C.C. Heyde, *Martingale Limit Theory and its Application*, Probability and Mathematical Statistics, Academic Press, New York-London-Toronto-Sydney-San Francisco (1980).

[31] Lars Peter Hansen, Large sample properties of generalized method of moments estimators, *Econometrica*, **50**, No. 4 (1982), 1029–1054.

[32] L. Peter Hansen, J. Scheinkman, Back to the future: generating moment implications for continuous-time Markov processes, *Econometrica*, **63**, No. 4 (1995), 767–804.

[33] R.Z. Has'Minskii, M.B. Nevel'Son, *Translations of Mathematical Monographs*, Stochastic Approximation and Recursive Estimation (Ed. B. Silver), Volume **47**, American Mathematical Society, Providence (1973); Translated from the Russian by Israel Program for Scientific Translations.

[34] H. He, *Moment Approximation and Estimation of Diffusion Models of Asset Prices*, Manuscript (1990).

[35] D. Heath, R. Jarrow, A. Morton, Bond pricing and the term structure of interest rates: a new methodology for contingent claims valuation, *Econometrica*, **60**, No. 1 (1992), 77–105.

[36] K. Itô, H. McKean, Jr., *Diffusion Processes and their Sample Paths*, Springer-Verlag, New York, Berlin (1974).

[37] C.S. Jones, *Bayesian Estimation of Continuous-time Finance Models*, Working Paper (1998).

[38] J.G. Kemeny, J.L. Snell, *Finite Markov Chains*, Undergraduate Text in Mathematics, Springer-Verlag, New York-Heidelberg-Berlin (1976).

[39] L.A. Klimko, P.I. Nelson, On conditional least squares estimation for stochastic processes, *Annals of Statistics*, **6** (1978), 629–642.

[40] P.R. Krugman, Target zones and exchange rate dynamics, *The Quarterly Journal of Economics*, **106**, No. 3 (1991), 669–682.

[41] H. Kunita, *Stochastic Flows and Stochastic Differential Equations*, Cambridge studies in advanced mathematics, Volume **24**, Cambridge University Press, Cambridge-New York-Melbourne (1990).

[42] H.J. Kushner, P.G. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuos Time*, Springer-Verlag, New York (1992).

[43] H.J. Kushner, D.S. Clark, Stochastic approximation methods for constrained and unconstrained systems, *Applied Mathematical Sciences*, **26**, Springer-Verlag, New York (1978).

[44] H.J. Kushner, G.G. Yin, *Stochastic Approximation Algorithms and Applications*, Applications of Mathematics, Volume **35**, Springer-Verlag, New York-Berlin-Heidelberg (1997).

[45] A.W. Lo, Maximum likelihood estimation of generalized Itô processes with discretely sampled data, *Econometric Theory*, **4** (1988), 231–247.

[46] T.A. Marsh, E.R. Rosenfeld, Stochastic processes for interest rates and equilibrium bond prices, *Journal of Finance*, **38** (1983), 635–646.

[47] M. Métivier, *Semimartingales, a Course on Stochastic Processes*, de Gruyter Studies in Mathematics, Volume **2**, Walter de Gruyter & Co., Berlin-New York (1982).

[48] G.N. Milstein, On the mean-square approximation of a diffusion process in a bounded domain, *Stochastics Stochastics Rep.*, **64**, No. 3-4 (1998), 211–233.

[49] G.N. Milstein, M.V. Tretyakov, Simulation of a space-time bounded diffusion, *Ann. Appl. Probab.*, **9**, No. 3 (1999), 732–779.

[50] D.B. Nelson, Arch models as diffusion approximations, *Journal of Econometrics*, **45** (1990), 7–38.

[51] L. Perko, *Differential Equations and Dynamical Systems*, Third Edition, Texts in Applied Mathematics, Volume **7**, Springer-Verlag, New York (2001).

[52] K. Petersen, *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, Volume **2**, Cambridge University Press, Cambridge-London-New York-New Rochelle-Melbourne-Sidney (1983).

[53] K. Singleton, Estimation of affine asset pricing models using the empirical characteristic function, *Journal of Econometrics* (2001).

[54] R.H. Stanton, A nonparametric model of term structure dynamics and the market price of interest rate of risk, *Journal of Finance*, **52** (1997), 1973–2002.

A. Appendix. Proofs

Proof of Theorem 1. If we define $T_n = \Theta_n - \theta^*$ then equation (11) becomes

$$T_{n+1} = T_n - \gamma_n H(\Theta_n, Y_n) V(\Theta_n, Y_n, X_{n+1}) \,.$$

It follows that

$$\begin{aligned}\mathbf{E}(\|T_{n+1}\|^2 \mid \mathcal{F}_n) - \|T_n\|^2 = & -2\gamma_n T_n \cdot H(\Theta_n, Y_n)\overline{V}(\Theta_n, Y_n) \\ & + \gamma_n^2 S^2(\Theta_n, Y_n) \leq \gamma_n^2 K(1 + \|T_n\|^2) \,, \end{aligned} \tag{35}$$

where the last inequality follows by equations (14) and (16). Moving the terms that have either $\|T_n\|$ or $\|T_{n+1}\|$ to the left of the previous equation we obtain

$$\mathbf{E}(\|T_{n+1}\|^2 \mid \mathcal{F}_n) - \|T_n\|^2(1 + K\gamma_n^2) \leq K\gamma_n^2 \,. \tag{36}$$

Define $\Pi_n = \prod_{k=1}^{n-1}(1 + K\gamma_n^2)$ and let $T'_n = (\frac{1}{\Pi_n})^{\frac{1}{2}} T_n$. We observe that the sequence (Π_n) is a convergent sequence of positive numbers since $(\log \Pi_n)$ converges by Hypothesis H_3. Using equation (36) we obtain

$$\mathbf{E}(\|T'_{n+1}\|^2 \mid \mathcal{F}_n) - \|T'_n\|^2 \leq K \frac{\gamma_n^2}{\Pi_{n+1}} \,. \tag{37}$$

If $F_n = \{\omega \in \Omega \mid \mathbf{E}(\|T'_{n+1}\|^2 \mid \mathcal{F}_n) - \|T'_n\|^2 \geq 0\}$, then equation (36) and Hypothesis H_3 imply that

$$\sum_{n=1}^{\infty} \mathbf{E}(\mathbf{1}_{F_n}(\|T'_{n+1}\|^2 - \|T'_n\|^2)) < \infty \,.$$

It follows by the almost sure convergence of quasi-martingales that $T_n^{'}$ converges almost surely toward a positive integrable random variable (see, for example, Theorem 9.4 page 49 and Proposition 9.5 of Métivier [47]). We conclude that the same property holds for T_n. The next step of the proof is to prove that the convergence of T_n is to zero. By inequality (37), Hypothesis H_3, the definition of T_n and the fact that Θ_1 belongs to $L^2(\mathbf{P})$, it follows that

$$\sup_n \mathbf{E}(\|T_n\|^2) < \infty \,. \tag{38}$$

We also observe that

$$\begin{aligned} 0 &\leq \sum_{n=1}^{\infty} 2\gamma_n \mathbf{E}(T_n \cdot H(\Theta_n, Y_n)\overline{V}(\Theta_n, Y_n)) \\ &\leq \sum_{n=1}^{\infty} \mathbf{E}(\|T_n\|^2) - \mathbf{E}(\|T_{n+1}\|^2) + (1 + \sup_{k \geq 0} \mathbf{E}(\|T_k\|^2)) \sum_{n=1}^{\infty} K\gamma_n^2 < \infty \,, \end{aligned}$$

where the last inequality follows by equations (35), (38), and Hypothesis H_3. Since $\sum_k \gamma_{n_k} = \infty$ there exists a subsequence (n'_k) of (n_k) such that

$$\lim_k \mathbf{E}(T_{n'_k} \cdot H(\Theta_{n'_k}, Y_{n'_k})\overline{V}(\Theta_{n'_k}, Y_{n'_k}) = 0\,, \tag{39}$$

equation (39) implies that for any $\varepsilon > 0$, $\liminf_k \|T_{n'_k}\| \le \varepsilon$ almost surely. To prove this last statement, let us assume otherwise. Then there exists $\varepsilon > 0$ such that $T_{n'_k} \ge \varepsilon$ for all k big enough on some set A of probability greater than zero. It would follow by Fubini Theorem and equation (15) that there exists $\delta > 0$ such that

$$\begin{aligned}\int_\Omega & T_{n'_k} \cdot H(\Theta_{n'_k}, Y_{n'_k})\overline{V}(\Theta_{n'_k}, Y_{n'_k}) d\mathbf{P} \\ &\ge \int_A T_{n'_k} \cdot H(\Theta_{n'_k}, Y_{n'_k})\overline{V}(\Theta_{n'_k}, Y_{n'_k})\, d\mathbf{P} \ge \int_A \delta d\mathbf{P} \ge \delta \mathbf{P}(A) > 0\end{aligned}$$

for all k big enough. The former is in contradiction with equation (39). Since (T_n) converges almost surely, then $T_n \to 0$ almost surely. □

Proof of Corollary 1. Let μ be a probability measure on $\mathbb{R}$ supported on S, and let $\mathbf{P} = \int \mathbf{P}_x^{\lambda^*} d\mu$. We define $\overline{V}\colon \mathbb{R} \times \mathtt{D} \mapsto \mathbb{R}$ and $S^2\colon \mathbb{R} \times \mathtt{D} \mapsto \mathbb{R}$ by the following formulas:

$$\overline{V}(\lambda, d) = \eta^f(\lambda^*, d) - \eta^f(\lambda, d)\,, \tag{40}$$

$$S^2(\lambda, d) = (\eta^f(\lambda^*, d) - \eta^f(\lambda, d))^2 + (\eta^{f^2}(\lambda^*, d) - (\eta^f(\lambda^*, d))^2)\,. \tag{41}$$

It follows by the strong Markov property of $(X_t, \mathcal{F}_t, \mathbf{P}_x^{\lambda^*})$ that Conditions A_1 and A_2 of Theorem 1 are satisfied. Since $\eta^f(\cdot, d)$ for any $d \in \mathtt{D}$ is bounded and $\mathtt{D}$ is finite it follows that property H_2 of Theorem 1 is satisfied. By Corollary 6 of Appendix B, equation (14) of Theorem 1 is satisfied. Last, we notice that

$$\begin{aligned}\mathbf{E}((\lambda - \lambda^*)&H(\lambda, X_{\tau_n})\overline{V}(\lambda, X_{\tau_n})) \\ &= \sum_{m=1}^{s} (\lambda - \lambda^*)\overline{V}(\lambda, d_m)\mathbf{P}(X_{\tau_n} = d_m)\,. \end{aligned} \tag{42}$$

By the last equation and Corollary 5 of Appendix A, equation (15) holds. It follows by Theorem 1 that the sequence of random variables (Θ_n) converges almost surely $\mathbf{P}$ to λ^*. Since the last statement holds for any initial probability measure supported on S the result follows. □

Proof of Corollary 2. Let μ be a probability measure on $\mathbb{R}$ supported on S, and let $\mathbf{P} = \int \mathbf{P}_x^{\varsigma^*}\, d\mu$. If we define $\widetilde{\widetilde{V}}\colon \mathbb{R}\times \mathtt{D} \mapsto \mathbb{R}$ and $\widetilde{S}^2\colon \mathbb{R}\times \mathtt{D} \mapsto \mathbb{R}$ by:

$$\widetilde{\widetilde{V}}(\varsigma, d) = \widetilde{\eta}^f(\varsigma^*, d) - \widetilde{\eta}^f(\varsigma, d)\,, \tag{43}$$

$$\widetilde{S}^2(\varsigma, d) = (\widetilde{\eta}^f(\varsigma^*, d) - \widetilde{\eta}^f(\varsigma, d))^2 + (\widetilde{\eta}^{f^2}(\varsigma^*, x) - (\widetilde{\eta}^f(\varsigma^*, x))^2)\,. \tag{44}$$

It follows by the strong Markov property of $(X_t, \mathcal{F}_t, \mathbf{P}_x^{\varsigma^*})$ that conditions A_1 and A_2 of Theorem 1 are satisfied. By assumptions 1 and 2 and Lemma 7 of Appendix B, it follows that Property H_2 of Theorem 1 is satisfied. The monotonicity and differentiability of σ_0, property 2 of Corollary 2 and Lemma 7 of Appendix B imply equation (14) of Theorem 1. Last, we notice that

$$\mathbf{E}((\varsigma - \varsigma^*)\widetilde{H}(\varsigma, X_{\tau_n})\widetilde{\widetilde{V}}(\varsigma, X_{\tau_n})) = \sum_{m=1}^{s} (\varsigma - \varsigma^*)\widetilde{\widetilde{V}}(\varsigma, d_m)\mathbf{P}(X_{\tau_n} = d_m)\,. \tag{45}$$

By the last equation and Corollary 5 of Appendix A, equation (15) holds. It follows by Theorem 1 that the sequence of random variables $(\widetilde{\Theta}_n)$ converges almost surely $\mathbf{P}$ to ς^*. Since the last statement holds for any initial probability measure supported on S the result follows. □

The following lemma is used in the proof of Theorem 2, Theorem 4 and Theorem 5.

Lemma 1. *Let $f\colon \mathtt{D} \to \mathbb{R}$ and $A = (a_{i,j})$ be the irreducible transition matrix of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ with left-fixed probability row vector $p = (p_i) >> 0$.*

Then for any $x \in S$

$$n^{-1}\sum_{k=1}^{n} f(X_{\tau_k}) \to \sum_{d\in \mathtt{D}} f(d)p_d \qquad \text{a.s. } \mathbf{P}_x\,. \tag{46}$$

Remark 1. A is a irreducible matrix. This follows by the recurrence of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$. Hence, there exists a unique left-fixed probability vector p; moreover, the entries of p are strictly positive (see for instance Petersen [52]).

Proof. Let $\tilde{\mathbf{P}}_d$ for $d \in \mathtt{D}$ be the restriction of $\mathbf{P}_d$ to $\sigma(X_{\tau_1}, X_{\tau_2}, \cdots)$ and let $\mathbf{P}$ be the probability measure defined on $\sigma(X_{\tau_1}, X_{\tau_2}, \cdots)$ by the formula $\mathbf{P} = \sum_{d\in\mathtt{D}} p_d \tilde{\mathbf{P}}_d$. By the strong Markov property θ_{τ_2} defines a measure-preserving transformation. Indeed, (X_{τ_n}) is an irreducible Markov shift. It follows by the point-wise ergodic theorem and the fact that an irreducible Markov shift is ergodic that equation (46) holds a.s. $\mathbf{P}$ (see Petersen [52]); the result follows using the strong Markov property and the fact that the components of the invariant vector p are positive. □

Theorem 5. *Let $Z\colon (\Omega, \mathcal{F}_{\tau_2}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a measurable map that is bounded below. Moreover, assume that $Z \in \bigcap_{d\in \mathtt{D}} L^2(\mathbf{P}_d)$ and $\mathbf{E}_d(Z) = 0$ for $d \in \mathtt{D}$. Let Z_n be defined as $Z_n = Z \circ \theta_{\tau_n}$, for $n \geq 2$ and $Z_1 = Z$. Then for any initial probability measure, the distribution of*

$$n^{-1/2} \sum_{k=1}^{n} Z_n\,, \tag{47}$$

approaches the normal distribution with mean zero and variance $\sigma^2 = \sum_{d\in\mathtt{D}} p_d \mathbf{E}_d(Z^2)$, where $p = (p_i)$ is the left-fixed probability row vector of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ as in Lemma 1.

Proof. Let μ be a probability measure on S, and let $(X_t, \mathcal{F}_t, \mathbf{P}_\mu)$ be the Markov process with initial probability measure μ. We observe that (Z_n) is a $(\mathcal{F}_{\tau_{n+1}})$ adapted process and

$$\mathbf{E}_\mu(Z_n \mid \mathcal{F}_{\tau_n}) = 0\,, \tag{48}$$

$$\mathbf{E}_\mu(Z_n^2) = \mathbf{E}_\mu \mathbf{E}_{X_{\tau_n}}(Z^2) < \infty\,, \tag{49}$$

for $n \geq 1$. The proof of the theorem follows using the strong Markov property, Lemma 1, and a line of argument similar to the technique of the proof of the Lindeberg-Lévy Theorem for martingales (see Billingsley [8]). □

For the proof of Theorem 3 we make use of the following easily proved lemma.

Lemma 2. *Let $\{\beta_{k,n} \mid 0 \leq k \leq n\}$ be the double indexed sequence of positive numbers defined as $\beta_{k,n} = \prod_{i=k+1}^{n}(1 - 1/i)$. Then*

$$(1 - \epsilon_k)\frac{k}{n} \leq \beta_{k,n} \leq (1 + \epsilon_k)\frac{k}{n}\,, \tag{50}$$

where (ϵ_k) is a sequence of positive numbers such that $\epsilon_k \to 0$ as $k \to \infty$. In particular $n^{1/2}\beta_{k,n} \to 0$ as $n \to \infty$ for any fixed k.

Proof of Theorem 3. We observe that the sequence of random variables (Θ_n^N) converges almost surely $\mathbf{P}$ to θ^* by Theorem 1. Let $\overline{V}\colon \mathbb{R} \times \mathtt{D} \to \mathbb{R}$ and

$S^2\colon \mathbb{R}\times \mathtt{D} \mapsto \mathbb{R}$ be defined by the formulas:

$$\overline{V}(\theta,d) = \eta(\theta^*,d) - \eta(\theta,d)\,,$$
$$S^2(\theta,d) = (\eta(\theta^*,d) - \eta(\theta,d))^2 + \mathbf{E}_d(Y - \eta(\theta^*,d))^2\,.$$

By Hypothesis N_1 it follows that for $\theta\in\mathbb{R}$, $d\in\mathtt{D}$,

$$(\theta-\theta^*)\frac{\overline{V}(\theta,d)}{(\partial\overline{V}/\partial\theta)(\theta^*,d)} \geq 0\,.$$

Hypothesis N_2 and the fact that α is defined on a finite set and is nowhere zero imply that S^2/α^2 satisfies Hypotheses H_2 of Theorem 1. Using the strong Markov property and an argument similar to the one used in the proof of Corollary 1 we can prove that (Θ_n^N) converges almost surely to θ^*. Let $Z\colon \mathtt{D}\times\mathbb{R}\to\mathbb{R}$ be defined as $Z(d,y) = y - \eta(\theta^*,d)$. We observe that $Z(d,y) = V(\theta,d,y) - \overline{V}(\theta,d)$ for any $\theta\in\mathbb{R}$. We denote as $\delta\colon \mathbb{R}\times\mathtt{D}\mapsto\mathbb{R}$ the function defined by the formula $\overline{V}(\theta,d) = (\partial/\partial\theta)\overline{V}(\theta^*,d)(\theta-\theta^*) + \delta(\theta-\theta^*,d)$. If we define $T_n^N = \Theta_n^N - \theta^*$ for $n\geq 1$, it follows that

$$T_{n+1}^N = (1-\frac{1}{n})T_n^N - \frac{Z_n}{n\alpha_n} - \frac{\delta_n}{n\alpha_n}\,, \tag{51}$$

where $Z_n = Z(X_{\tau_n},Y_{n+1})$, $\delta_n = \delta(T_n^N, X_{\tau_n})$ and $\alpha_n = \alpha(X_{\tau_n})$. Iteration of equation (51) yields

$$T_{n+1}^N = \beta_{0,n}T_1^N - \sum_{k=1}^n \frac{\beta_{k,n}}{k}\frac{\delta_k}{\alpha_k} - \sum_{k=1}^n \frac{\beta_{k,n}}{k}\frac{Z_k}{\alpha_k}\,. \tag{52}$$

Hence, we can prove that $n^{1/2}T_n^N$ is asymptotically normal with mean zero and variance σ^2, by proving that

$$n^{1/2}\beta_{0,n}T_1^N \to 0 \quad \text{almost surely,} \tag{52a}$$
$$n^{1/2}\sum_{k=1}^n \frac{\beta_{k,n}}{k}\frac{\delta_k}{\alpha_k} \to 0 \quad \text{in probability,} \tag{52b}$$
$$n^{1/2}\sum_{k=1}^n (\frac{\beta_{k,n}}{k} - \frac{1}{n})(\frac{Z(X_{\tau_k},Y_{k+1})}{\alpha_k}) \to 0 \quad \text{in probability,} \tag{52c}$$
$$n^{-1/2}\sum_{k=1}^n \frac{Z(X_{\tau_k},Y_{k+1})}{\alpha_k} \to N(0,\sigma^2) \quad \text{in distribution.} \tag{52d}$$

Equation (52a) follows by Lemma 2. Next, we observe that the terms on the left hand side of equation (52c) are uncorrelated. We observe that $Y \in \bigcap_{d\in\mathtt{D}} L^2(\mathbf{P}_d)$ and α is a nowhere zero function defined on a finite set. It follows by the strong Markov property, and Lemma 2 that there exists a constant $C > 0$ such that

$$\mathbf{E}(n^{1/2}\sum_{k=1}^{n}(\frac{\beta_{k,n}}{k} - \frac{1}{n})(\frac{Z(X_{\tau_k}, Y_{k+1})}{\alpha(X_{\tau_k})}))^2 \leq \frac{C}{n}\sum_{k=1}^{n}\epsilon_k^2\,, \tag{53}$$

where the right-hand side of equation (53) goes to zero as $n \to \infty$ since $\epsilon_k \to 0$ as $k \to \infty$. Convergence in equation (52c) follows by Chebyshev inequality. Next we prove the convergence of equation (52b). We observe that equation (37) and the proof of Theorem 1 imply

$$\limsup_n n\,\mathbf{E}(T_n^N)^2 < \infty\,. \tag{54}$$

Let $\epsilon_1, \epsilon_2 > 0$, and let $\overline{\delta}(\cdot) = \max_{d\in\mathtt{D}}(\delta(\cdot, d)/\alpha(d))$. Since $\overline{\delta}(x) = o(\mid x \mid)$ there exists $\epsilon' > 0$ such that

$$\mid \overline{\delta}(x) \mid \leq \epsilon_2^2 \mid x \mid \qquad \text{for} \ \mid x \mid \leq \epsilon'\,. \tag{55}$$

Since $T_n^N \to 0$ almost surely, there exist $N_1 > 0$ such that

$$\mathbf{P}(\mid T_k^N \mid \leq \epsilon', k \geq N_1) > 1 - \epsilon_1\,. \tag{56}$$

It follows using equation (56), the triangle inequality, equation (55), Markov inequality, and Lyapounov inequality that

$$\begin{aligned}
&\mathbf{P}(\mid n^{1/2}\sum_{k=N_1}^{n}\frac{\beta_{k,n}}{k}\frac{\delta_k}{\alpha_k} \mid > \epsilon_2) \\
&\leq \epsilon_1 + \mathbf{P}(\mid n^{1/2}\sum_{k=N_1}^{n}\frac{\beta_{k,n}}{k}\frac{\delta_k}{\alpha_k} \mid > \epsilon_2, \mid T_k^N \mid \leq \epsilon', k \geq N_1) \\
&\leq \epsilon_1 + \mathbf{P}(n^{1/2}\sum_{k=N_1}^{n}\frac{\beta_{k,n}}{k} \mid \frac{\delta_k}{\alpha_k} \mid > \epsilon_2 \mid, T_k^N \mid \leq \epsilon', k \geq N_1) \\
&\leq \epsilon_1 + \mathbf{P}(\epsilon_2^2 n^{1/2}\sum_{k=N_1}^{n}\frac{\beta_{k,n}}{k} \mid T_k^N \mid > \epsilon_2) \\
&\leq \epsilon_1 + \epsilon_2\,\mathbf{E}(n^{1/2}\sum_{k=N_1}^{n}\frac{\beta_{k,n}}{k} \mid T_k^N \mid) \\
&\leq \epsilon_1 + \epsilon_2(\sum_{k=N_1}^{n}\frac{\beta_{k,n}}{k} n^{1/2}\mathbf{E}^{1/2}(T_k^N)^2)\,.
\end{aligned} \tag{57}$$

The convergence in probability in equation (52b) follows using equation (57) and Lemma 2. Finally, we observe that the convergence in distribution of equation (52d) is a consequence of Theorem 5. □

Proof of Corollary 3. We observe that $X_{\nu_n} = X_{\nu_2} \circ \theta_{\tau_{n-1}} = X_{\nu_1} \circ \theta_{\tau_2}^{n-2}$ for $n \geq 3$. Indeed, η satisfies condition N_1 of Theorem 3 by Corollary 6, and by the definition of η condition N_2 is satisfied. The result is a straightforward consequence of Theorem 3. □

Proof of Corollary 4. We observe that $\nu_n - \tau_{n-1} = \nu_2 \circ \theta_{\tau_{n-1}} = \nu_2 \circ \theta_{\tau_2}^{n-2}$ for $n \geq 3$. By Lemma 7 $\tilde{\eta}$ satisfies condition N_1 of Theorem 3, and $\tilde{\eta}$ satisfies condition N_2 by Lemma 7. The result is a straightforward consequence of Theorem 3. □

B. Appendix

In this appendix we derive some technical results about the transition matrices of the Markov chain $(X_{\tau_n}, \mathcal{F}_{\tau_n})$. In this section all the matrices are stochastic matrices. We state some easily proved results. The proof is left to the reader.

Definition 1. We say that a matrix A of size $s \times s$ is of type **I** if for all $i, j \in \{1 \dots s\}$, $i \equiv j \mod s$ implies $a_{i,j} = 0$. We would say that a matrix A of size $s \times s$ is of type **II** if whenever $i \equiv j + 1 \mod s$ implies $a_{i,j} = 0$ for all $i, j \in \{1 \dots s\}$.

Lemma 3. *Let A and B be two $s \times s$ matrices. Then the following holds:*

If A and B are both matrices of type **I** *then AB is a matrix of type* **II**.

If A and B are matrices of type **II** *then so is AB.*

If A is a matrix of type **I** *and B is a matrix of type* **II** *then AB and BA are matrices of type* **I**.

Definition 2. Given a matrix A of type **II** we define $P_{even}(A)$ and $P_{odd}(A)$ to be the matrices of size $[\frac{s}{2}] \times [\frac{s}{2}]$, and size $[\frac{s+1}{2}] \times [\frac{s+1}{2}]$ respectively defined by the following formulas

$$(P_{even}(A))_{i,j} = a_{2i,2j} \qquad \text{for } i, j \in \{1 \dots [\frac{s}{2}]\},$$
$$(P_{odd}(A))_{i,j} = a_{2i-1,2j-1} \quad \text{for } i, j \in \{1 \dots [\frac{s+1}{2}]\}.$$

Lemma 4. *If A and B are $s \times s$ matrices of type* **II** *then the following property holds:*

$$P_{even}(A)P_{even}(B) = P_{even}(AB)\,,$$
$$P_{odd}(A)P_{odd}(A) = P_{odd}(AB)\,.$$

In particular for any n positive integer

$$P_{even}(A^n) = (P_{even}(A))^n\,,$$
$$P_{odd}(A^n) = (P_{odd}(A))^n\,.$$

It is obvious that a matrix A of type **II** is completely determined by $P_{odd}(A)$ and $P_{even}(A)$.

Lemma 5. *Let $A = (a_{i,j})$ be an $s \times s$ matrix, whose entries are non-negative and such that $a_{i,j} \neq 0$ for $\mid i-j \mid \leq 1$. Then for any positive integer n, A^n is a matrix such that $a^n_{i,j} \neq 0$ for $\mid i-j \mid \leq n$, where $A^n = (a^n_{i,j})_{i,j}$*

We observe that if A is the transition matrix of the Markov process $(X_{\tau_n}, \mathcal{F}_{\tau_n})$ then $P_{odd}(A^2)$ and $P_{even}(A^2)$ satisfies the condition of the previous lemma.

Corollary 5. *If A is an $s \times s$ transition matrix of a Markov process $(X_{\tau_n}, \mathcal{F}_{\tau_n})$, then $P_{odd}(A^{2n})$ and $P_{even}(A^{2n})$ converge to stochastic matrices A_1 and A_2. Indeed, there are matrices C_1 and C_2, where C_1 is a $1 \times [\frac{s+1}{2}]$ matrix and C_2 is a $1 \times [\frac{s}{2}]$ matrix such that $A_1 = (1,\dots,1)'C_1$ and $A_2 = (1,\dots,1)'C_2$ and the components of C_1 and C_2 are positive.*

Proof. The result follows from Lemma 4, Lemma 5 , the observation made right after the proof of Lemma 5 and the fundamental theorem for regular Markov chains (see for example Kemeny and Snell [38], Theorem 4.1.4). □

C. Appendix

In this appendix we state and prove some results that are needed for the actual computation of the moments required for the construction of the algorithms proposed. Let us assume that $(X_t, \mathcal{F}_t, \mathbf{P}_x)$ is a regular diffusion on S, where S is a interval of $\mathbb{R}$. We assume that the differential operator L of the diffusion is given by

$$L = \frac{1}{2}\sigma^2(x)\frac{d^2}{dx^2} + b(x)\frac{d}{dx}\,, \tag{58}$$

where $\sigma^2 \colon S \mapsto \mathbb{R}^+$, $b \colon S \mapsto \mathbb{R}$ satisfy condition 2. Moreover we assume that $b/\sigma^2 \in C([c,d]) \cap C^2((c,d)$, where $c, d \in S$ and $c < d$. It follows that $s \colon [c,d] \to \mathbb{R}$ defined by

$$s(x) = \int_c^x \exp\{-\int_c^y \frac{2b(z)}{\sigma^2(z)}\, dz\}\, dy\,, \tag{59}$$

belongs to $C([c,d]) \cap C^2((c,d))$. It is an elementary exercise to verify that s satisfies the equation $Ls = 0$, with initial condition $s(c) = 0$. It follows by Theorem 13.16 volume II of Dynkin [19] that

$$\{f(d) - f(c)\}\frac{s(x)}{s(d)} + f(c) = \mathbf{E}_x f(X_{\tau_c \wedge \tau_d}))\,. \tag{60}$$

Let Λ be an interval of $\mathbb{R}$. Assume that $(X_t, \mathcal{F}_t, \mathbf{P}_x^\lambda)_{\lambda \in \Lambda}$ is a parametric set of diffusions on $\mathbb{R}$, with sample space $(\Omega, \mathcal{F}_\infty)$ and differential operators $(L_\lambda)_{\lambda \in \Lambda}$, where L_λ is given by the formula

$$L_\lambda = \frac{1}{2}\sigma^2(x,\lambda)\frac{d^2}{dx^2} + b(x,\lambda)\frac{d}{dx}\,. \tag{61}$$

Here we assume that $b(\cdot,\lambda)$ and $\sigma^2(\cdot,\lambda)$ satisfy the hypotheses of this appendix where as before $c < d$ belongs to S are fixed constants. We wish to find conditions on b, σ^2 to guarantee that the function given by $\lambda \mapsto \mathbf{E}_x^\lambda f(X_{\tau_c \wedge \tau_d})$ is monotone decreasing (or increasing). For this we prove the following lemma.

Lemma 6. *Let $c < d$ be real numbers and let Λ be a closed interval of $\mathbb{R}$. Let $f \colon [c,d] \times \Lambda \mapsto (0,\infty)$ be a jointly continuous positive function such that $\partial f/\partial\lambda$ is also jointly continuous. Let us assume that*

$$\frac{\partial f}{\partial \lambda}(x,\lambda) = f(x,\lambda) g(x,\lambda)\,,$$

where g is a strictly increasing (strictly decreasing) function in x for each $\lambda \in \Lambda$. It follows that the function

$$h(x,\lambda) = \frac{\int_x^d f(y,\lambda)\, dy}{\int_c^x f(y,\lambda)\, dy}$$

is a strictly increasing (strictly decreasing) function in Λ for any $x \in [c,d]$.

Proof. We prove g strictly increasing in x implies that h is a strictly increasing function in λ (the proof of g strictly decreasing implies that h is strictly decreasing is similar). By the dominated convergence theorem

$$\frac{\partial h}{\partial \lambda}(x,\lambda) = \frac{\int_c^x f(y,\lambda)\, dy \int_x^d \frac{\partial f}{\partial \lambda}(y,\lambda)\, dy - \int_x^d f(y,\lambda)\, dy \int_c^x \frac{\partial f}{\partial \lambda}(y,\lambda)\, dy}{\{\int_c^x f(y,\lambda)\, dy\}^2}\,.$$

The result follows using the following inequalities:

$$(\int_x^d f(y,\lambda)\,dy)g(x,\lambda) < \int_x^d \frac{\partial f}{\partial\lambda}(y,\lambda)\,dy < (\int_x^d f(y,\lambda)\,dy)g(d,\lambda)\,,$$

$$(\int_c^x f(y,\lambda)\,dy)g(c,\lambda) < \int_c^x \frac{\partial f}{\partial\lambda}(y,\lambda)\,dy < (\int_c^x f(y,\lambda)\,dy)g(x,\lambda)\,. \qquad \square$$

Corollary 6. *Let $(X_t,\mathcal{F}_t,\mathbf{P}_x^\lambda)_{\lambda\in\Lambda}$ be a parametric set of diffusions with differential operators $(L_\lambda)_{\lambda\in\Lambda}$ as in equation (61). Assume that $b(\cdot,\lambda)$ and $\sigma^2(\cdot,\lambda)$ satisfy the hypotheses of this appendix, where as before $c<d$ in S are fixed constants. Let $\eta\colon [c,d]\times\Lambda\mapsto\mathbb{R}$ be defined by the formula $\eta(x,\lambda)=\mathbf{E}_x^\lambda f(X_{\tau_c\wedge\tau_d})$. If $\partial/\partial\lambda(b/\sigma^2)(x,\lambda)>(<)0$ for all $(x,\lambda)\in[c,d]\times\Lambda$ then $\eta(x,\lambda)$ is a strictly decreasing (increasing) function on λ for all $x\in[c,d]$*

Proof. The result follows by Lemma 6 and equation (61). $\square$

Finally, we mention a result that allows us, in the case of diffusions with an one-dimensional state space, to compute the expected values of exit times from open sets.

Lemma 7. *Let $(X_t,\mathcal{F}_t,\mathbf{P}_x)$ be a regular diffusion on S, where S is an interval of $\mathbb{R}$. We assume that L is the differential operator of the diffusion where L is defined by equation (58) and we assume that σ^2 and b satisfy the hypothesis of this appendix. Set*

$$\varphi(x)=\exp\{-\int_c^x \frac{2b(z)}{\sigma^2(z)}\,dz\}\,.$$

Then

$$\mathbf{E}_x\tau_c\wedge\tau_d=\eta(x)=-\int_c^x 2\varphi(y)\int_c^y \frac{1}{\sigma^2(z)\varphi(z)}\,dz\,dy + \frac{\int_c^x\varphi(z)\,dz}{\int_c^d\varphi(z)dz}\int_c^d 2\varphi(y)\int_c^y\frac{1}{\sigma^2(z)\varphi(z)}\,dz\,dy\,. \quad (62)$$

Moreover, $u(x)=\mathbf{E}_x(\tau_c\wedge\tau_d)^2<\infty$ for any $x\in[c,d]$ and u is the solution of the differential equation

$$Lu=-\eta\,, \quad (63)$$

with boundary conditions $u(c)=u(d)=0$.

Proof. Equation (62) follows by Theorem 13.16 volume II of Dynkin [19] and a straightforward computation. The later part of the lemma follows by Theorem I.15.3 of Gihman and A.V. Skorohod [28]. $\square$