WORKING PAPER

No: 12-05
October, 2012

Estimating Stochastic Differential Equations Using Repeated Eigenfunction Estimation and Neural Networks

Ruhi Tuncer

giam
GALATASARAY UNIVERSITY
ECONOMIC RESEARCH CENTER

Estimating Stochastic Differential Equations Using Repeated Eigenfunction Estimation and Neural Networks

Ruhi Tuncer
Department of Economics
Galatasaray University
Istanbul Turkey

October 24, 2012

Abstract

We propose identifying the drift and the diffusion functions of an ergodic scalar stochastic differential equation using repeated eigenfunction estimation. The transition density will be estimated in a new way involving Kolmogorov's backward equation, neural networks and functions of our choice. Martingale estimating functions will be used to obtain asymptotic properties.

Keywords Stochastic Differential Equation, Kolmogorov's backward Equation, Infinitesimal Generator, Eigenfunctions, Transition Density, Neural Networks, Martingale Estimating Functions.

1 Introduction

Interest has been focused on the identification of the drift and diffusion functions of stationary stochastic differential equations. Emphasis has been placed on the infinitesimal generator of the underlying process. The infinitesimal generator, as an operator on a Hilbert space has eigenvalues and eigenfunctions which along with a nonparametric estimator of the stationary density, can be used to find the local variance function. In this paper, following a review of the framework presented in the seminal article by Hansen and Scheinkman (1995), we will propose an extension of these techniques. We will consider repeated eigenfunction estimation as an alternative.

We have the stochastic differential equation

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{1}$$

where $\mu(X_t)$ is the drift and $\sigma(X_t)$ is the diffusion. Let $\phi(X_t)$ be a twice continuously differentiable function with $E[\phi(X_t)] = 0$ and $E[\phi^2(X_t)] < \infty$ where the expected values are taken with respect to the stationary distribution $q(X_t)$ of X_t.

Let

$$A\phi = lim_{t\to 0}\frac{E[\phi(X_t)]X_0] - \phi(X_0)}{t}$$

be the infinitesimal generator of the process X_t. The infinitesimal generator satisfies Kolmogorov's backward equation:

$$\frac{\partial\phi(t,x)}{\partial t} = \mu(x)\frac{\partial\phi(t,x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2\phi(t,x)}{\partial x^2}. \tag{2}$$

This partial differential equation can be solved using separation of variables.

Let $\phi(t,x) = \phi(x)\psi(t)$.

$$\psi'(t)\phi(x) = \mu\psi(t)\phi'(x) + \frac{1}{2}\sigma^2(x)\phi''(x)\psi(t)$$

i.e.

$$\frac{\psi'(t)}{\psi(t)} = \frac{\mu\phi'(x)}{\phi(x)} + \frac{1}{2\phi(x)}\sigma^2(x)\phi''(x).$$

Since the left hand side is a function of t and the right hand side is a function of x these side are equal to a constant λ. So

$$\frac{\psi'(t)}{\psi(t)} = \lambda$$

and

$$\frac{\mu\phi'(x)}{\phi(x)} + \frac{1}{2\phi(x)}\sigma^2(x)\phi''(x) = \lambda.$$

We have

$$\psi = exp(\lambda t)$$

and

$$\mu(x)\frac{\partial\phi(x)}{\partial x}+\frac{1}{2}\sigma^2(x)\frac{\partial^2\phi(x)}{\partial x^2}=\lambda\phi(x).$$

So that

$$\phi(t,x)=exp(\lambda t)\phi(x).$$

On the other hand we know that the solution of Kolmogorov's backward equation is given by

$$\phi(t,x)=E[\phi(X_t)|X_0=x].$$

So that we have

$$E[\phi(X_t)|X_0=x]=exp(\lambda t)\phi(x).$$

Obviously

$$E[\phi(X_{t+s})|X_t]=exp(\lambda s)\phi(X_t). \tag{3}$$

Kessler and Sorensen (1996) were the first to use this relation to identify eigenfunctions. In section 3, we will use this relation to estimate the eigenfunctions in a new way.

The infinitesimal generator of a diffusion X_t on a compact interval $[l,r]$ with two reflective barriers and a strictly positive diffusion coefficient $\sigma^2(X_t)$ is self adjoint and negative semidefinite so that the standard Sturm-Liouville theory applies.

$A\phi=-\lambda\phi$ with $\lambda>0$ and $\phi'(l)=\phi'(r)=0$, a twice continuously differentiable solution to this eigenvalue problem will result in an eigenfunction for A.

Taking the expected value of the infinitesimal generator with respect to the stationary distribution $q(X_t)$ we have

$$E[A\phi]=0.$$

In other words

$$\int_R A\phi q dx=0.$$

The infinitesimal generator of a diffusion X_t with reflective barriers $[l,r]$ where ϕ is continuous and twice continuously differentiable with $\phi'(l)=0$ and $\phi'(r)=0$ can be expressed in terms of the stationary distribution. We have

$$\int_{[l,r]}[\mu\phi'+\frac{1}{2}(\sigma^2]\phi'']q=0.$$

Integration by parts gives us

$$\int_{[l,r]} [\mu q - \frac{1}{2}(\sigma^2 q)']\phi' + \frac{1}{2}(\sigma^2(x) q \phi'|_l^r.$$

Since $\phi'(l) = \phi'(r) = 0$, we have

$$\int_{[l,r]} [\mu q - \frac{1}{2}(\sigma^2 q)']\phi' = 0.$$

We can choose ϕ in a way that we have

$$[\mu q - \frac{1}{2}(\sigma^2 q)'] = 0$$

Now

$$\mu \phi' q + \frac{1}{2}\sigma^2 \phi'' q = \lambda \phi q$$

so that we have

$$A\phi = \frac{(\sigma^2 q \phi')'}{2q} = \lambda \phi q.$$

and

$$\sigma^2(x) = \frac{2\lambda \int_l^x \phi q dx}{q\phi'} \tag{4}$$

(Hansen and Scheinkman (1995) and Hansen, Scheinkman, Ait-Sahalia (2010)).

Hansen, Scheinkman and Touzi (1998) have proved that the eigenfunction which corresponds to the eigenvalue closest to zero is monotonic. They have also shown that $lim_{l \to 0} q(l)\phi'(l) = 0$.

Ait-Sahalia (1996) has proposed a method to estimate the diffusion coefficient of a stochastic differential equation having an affine drift.

Demoura (1998) has proposed estimating two eigenfunctions in order to estimate the drift and the diffusion functions simultaneously.

Let ϕ_1 and ϕ_2 be two eigenfunctions associated with eigenvalues λ_1 and λ_2 .

$$\mu \phi_1' + 1/2 \sigma^2 \phi_1'' = \lambda_1 \phi_1$$

and

$$\mu \phi_2' + 1/2 \sigma^2 \phi_2'' = \lambda_2 \phi_2.$$

The simultaneous solution of these equations gives μ and σ.

In section 2, we will review neural networks. In section 3, we will present a new way of identifying the drift and diffusion functions. In section 4, we will present a new way of estimating the transition density and the eigenfunctions. In section 5, we will study the asymptotic properties of the estimators we propose.

2 Review of Derivative estimation Using Artificial Neural Networks

Eigenfuncions of the infinitesimal generator are unknown so that we have to use certain techniques to approximate them. A number of authors have shown that feedforward neural networks are capable of approximating a large class of functions and their partial derivatives (Cybenko (1989), Hornik. Stinchombe and White (1990) and Hornik (1991)). Let ψ be the common activation function, a , β and γ vectors of parameters and x a vector of variables.

Let $\mathcal{N}_k^{(n)}(\psi) = \{e : R^k \to R | e(x) = \sum_{j=1}^{n} \beta_j \psi(a_j' x - \gamma_j)\}$ be the set of functions implemented by a neural network with n hidden units and a single output unit.

$\mathcal{N}_k(\psi) = \bigcup_{n=1}^{\infty} \mathcal{N}_k^{(n)}(\psi)$ is the set of functions implemented by a network with any number of hidden units.

Let $\alpha = (\alpha_1, ..., \alpha_k)$ of non-negative integers be a multi-index. $|\alpha| = \alpha_1 + ... + \alpha_k$ is the order of the multi-index.

$$D^\alpha f(x) = \frac{\partial^{\alpha_1 + ... + \alpha_k} f}{\partial x_1^{\alpha_1} ... \partial x_k^{\alpha_k}}$$

is the partial derivative of a function f of $x = (x_1, ..., x_k)$.

Let $\mathcal{C}^m(R^k)$ be the space continuous functions with continuous partial derivatives of order $|\alpha| \leq m$. Let Q be a finite measure on R^k. For $f \in \mathcal{C}^m(R^k)$ the following norm is defined:

$$||f||_{m,p,Q} = \{ \sum_{|\alpha| \leq m} \int_{R^k} |D^\alpha f|^p dQ \}^{1/p}$$

for $1 \leq p < \infty$.

The weighted Sobolev space $\mathcal{C}^{m,p}(Q)$ is defined as follows:

$$\mathcal{C}^{m,p}(Q) = \{ f \in \mathcal{C}^m(R^k) : ||f||_{m,p,Q} \leq \infty \}$$

.

A subset S of $\mathcal{C}^{m,p}(Q)$ is dense in $\mathcal{C}^{m,p}(Q)$ if for all $f \in \mathcal{C}^{m,p}(Q)$ there exists a function $g \in S$ such that $||f-g||_{m,p,Q} < \epsilon$ for all $\epsilon > 0$.

We are now ready to state Hornik's theorem.

Theorem (Hornik 1991 Theorem 4): If $\psi \in \mathcal{C}^m(R^k)$ is non-constant and all of its derivatives up to order m are bounded, then $\mathcal{N}_k^{(n)}(\psi)$ is dense in $\mathcal{C}^{m,p}(Q)$ for all finite measures Q on R^k .

This result is the basis for the neural network approximations that we will make. In the literature interest has been on approximating derivatives. In the following, we will have to deal with integrals of neural networks approximations, so that we need the following property:

Lemma Let $a = (a_1, ..., a_k)$. If $||f-g||_{m,p,Q} < \epsilon$ for all $\epsilon > 0$ then

$$|| \int_a^x f dQ - \int_a^x g dQ ||_{m,p,Q} < \epsilon.$$

Proof

$$|| \int_a^x f dQ - \int_a^x g dQ ||_{m,p,Q} < \int_{R^k} ||f-g||_{m,p,Q} d\mu.$$

Choosing $||f-g||_{m,p,Q} < \epsilon/Q$ does the trick.

Q.E.D.

3 Identification of Scalar Diffusions Using Repeated Eigenfunction Estimation

In this section we propose a new way of expressing the diffusion function of a stochastic differential equation We have

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t.$$

We observe that if the eigenfunction $\phi(x_t)$ of the infinitesimal generator

$$A\phi = \mu(x)\frac{\partial \phi(x)}{\partial x} + \frac{1}{2}\sigma^2(x)\frac{\partial^2 \phi(x)}{\partial x^2}$$

is a twice continuously differentiable function, then by Ito's Lemma it follows the stochastic differential equation

$$d\phi(X_t) = A\phi(X_t)dt + \sigma(X_t)\frac{\partial \phi(x)}{\partial x}dW_t.$$

To simplify the notation we will write ϕ_t for $\phi(X_t)$ and do the same for the other functions. We have

$$d\phi_t = -\lambda_1 \phi_t dt + \sigma_t \frac{\partial \phi(x)}{\partial x} dW_t.$$

Using the technique summarized in section 3 , $\phi(x_t)$ can be estimated and

$$\frac{\partial \phi(x)}{\partial x}$$

can be calculated analytically. σ_t is the only unknown in this equation.Now the data we have are x_t . Once the eigenfunction $\phi(x_t)$ is estimated we can generate the series ϕ_t and try to estimate the drift and diffusion functions of the equation

$$d\phi_t = -\lambda_1 \phi_t dt + \sigma_t \frac{\partial \phi(x)}{\partial x} dW_t.$$

This equation also has an infinitesimal generator whose eigenfunctions can be estimated using the procedure we have already used. Let L be the infinitesimal generator of the equation satisfied by ϕ_t and f be its eigenfunction. We have

$$Lf(\phi_t) = -\lambda_1 \phi_t \frac{\partial f}{\partial \phi} + \frac{1}{2} \sigma_t^2 [\phi_t']^2 \frac{\partial^2 f}{\partial \phi^2}.$$

$$Lf(\phi_t) = -\lambda_2 f(\phi_t)$$

so that

$$-\lambda_1 \phi_t \frac{\partial f}{\partial \phi} + \frac{1}{2} \sigma_t^2 [\phi_t']^2 \frac{\partial^2 f}{\partial \phi^2} = -\lambda_2 f(\phi_t)$$

i.e.

$$\sigma_t^2 = \frac{2}{[\phi_t']^2 f''} [-\lambda_2 f(\phi_t) + \lambda_1 \phi_t f'] \tag{5}$$

where

$$f' = \frac{\partial f}{\partial \phi}, f'' = \frac{\partial^2 f}{\partial \phi^2}.$$

To find $\mu(X_t)$, we can use the relation

$$\mu(x) \frac{\partial \phi(x)}{\partial x} + \frac{1}{2} \sigma^2(x) \frac{\partial^2 \phi(x)}{\partial x^2} = \lambda \phi(x).$$

4 Estimating the Transition Density and the Eigenfunctions

To estimate the eigenfunctions of the infinitesimal generator, we will have to estimate the transition density of the process. Ait-Sahalia (2002) has used Hermite expansions to estimate the transition density. Here we propose a different method. Let

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t$$

be a time homogeneous stochastic differential equation. Let $h(X_t) \in C_0^2$ be a continuous function with compact support, having continuous derivatives of order two.

Suppose that h is a function of our choice and has the following properties: $E[h(X_t)]$ is independent of time t and $E[h^2] < \infty$.

Let $p(X_t, X_{t-1})$ be the time homogeneous transition density of the process.

Let $p(X_t, X_{t-1}, \gamma)$ be a neural networks approximation of $p(X_t, X_{t-1})$.

The function $u(X_{t-1}) = E[h(X_t)|X_{t-1}]$ satisfies Kolmogorov's backward equation.

Now $E[h(X_t)|X_{t-1}] = \int_D h(y)p(y, X_{t-1})dy$. A neural networks approximation of this expression will be

$$u(X_{t-1}, \gamma) = \int_D h(y)p(y, X_{t-1}, \gamma)dy \tag{6}$$

Notice that the integral will be evaluated analytically. This aspect of the procedure can be simplified by a judicious choice of the function h and the neural networks approximation $p(X_t, X_{t-1}, \gamma)$.

Now $E|u(X_{t-1}) - h(X_t)|X_{t-1}] = 0$.

So that we have a martingale difference and we can use martingale estimation techniques. Since we have $E[u(X_{t-1}, \gamma)] = E[h(X_t)]$ the sample analog will be

$$\frac{1}{T}\sum_{t=0}^{T}[u(X_{t-1}, \gamma) - h(X_t)]. \tag{7}$$

The resulting estimate of the transition density will be denoted $p(x, y, \hat{\gamma})$.

Let $\phi(x, \theta)$ be a neural networks approximation of $\phi(x)$. Then the conditional expected value can be approximated by

$$\int_D \phi(y,\theta)p(y,X_{t-1},\hat{\gamma})dy.$$

Since

$$E[[\phi(X_t)|X_{t-1})] - exp(\lambda)\phi(X_{t-1})|X_{t-1}] = 0$$

its sample analog will be

$$\frac{1}{T}\sum_{t=0}^{T}[\int_D \phi(y,\theta)p(y,X_{t-1},\hat{\gamma})dy - exp(\lambda)\phi(X_{t-1})].$$

On the other hand, since

$$E[E[\phi(X_t)|X_{t-1})] - exp(\lambda)\phi(X_{t-1})] = 0$$

another sample analog will be

$$\frac{1}{T}\sum_{t=0}^{T}[\int_D \phi(y,\theta)p(y,X_{t-1},\hat{\gamma})dy - exp(\lambda)\phi(X_t)]. \qquad (8)$$

4.1 Activation Functions

To approximate $\phi(X_t)$ using neural networks, we will have to choose activation functions. A combination of logistic functions has been fruitful in empirical studies. Let α and β be vectors of parameters. Let ψ be the following activation function.

$$\psi_i(x) = \frac{1}{1+e^{-\alpha_i x}}.$$

$$\phi(x,\theta) = \sum_{i=1}^{n}\beta_i\psi_1(\alpha_i x) \qquad (9)$$

is the function we will use to approximate $\phi(X_t)$.

Since the cumulative distribution function is sigmoid, we will specify the transition density using a gaussian radial basis function.

Let $\gamma = (a_i, b_i, c_i, d_i, e_i)$ be the parameter vector. Let

$$\rho(x-c_i) = exp[-b_i(x-c_i)^2]$$

and

$$\eta(y - e_i) = exp[-d_i(y - e_i)^2]$$

$\psi(x, y, \gamma)$ is the activation function given by

$$\psi(x, y, \gamma) = \frac{\sum_{i=1}^{n} \rho(x - c_i)\eta(y - e_i)}{\sum_{i=1}^{n} \rho(x - c_i)}$$

i.e.

$$p(x, y, \gamma) = \frac{\sum_{i=1}^{n} \rho(x - c_i)\eta(y - e_i)}{\sum_{i=1}^{n} \rho(x - c_i)}$$

The function $h(y)$ to be chosen to estimate the transition density can be of the form

$$h(y) = y \tag{10}$$

so that

$$u(X_{t-1}, \gamma) = \int_D h(y)p(y, X_{t-1}, \gamma)dy = \frac{\sum_{i=1}^{n} \rho(x - c_i) \int_D y\eta(y - e_i)dy}{\sum_{i=1}^{n} \rho(x - c_i)} \tag{11}$$

facilitating integration. This is like estimating the transition density from the conditional expected value. Different choices of h are possible. The expression

$$\frac{1}{T}\sum_{t=0}^{T}[u(X_{t-1}, \gamma) - h(X_t)]$$

will then become

$$\frac{1}{T}\sum_{t=0}^{T}[\frac{\sum_{i=1}^{n} \rho(X_{t-1} - c_i) \int_D y\eta(y - e_i)dy}{\sum_{i=1}^{n} \rho(X_{t-1} - c_i)} - X_t]. \tag{12}$$

Despite the complicated appearance of $u(X_{t-1}, \gamma)$, this expression will reduce to linear combinations of gaussian densities in both the numerator and the denominator.

Notice that our results do not depend on the particular neural networks activation functions that we have chosen. Any activation function will do as long as the conditions of theorems 2.2 and 2.3 of Sorensen (2009) are satisfied.

5 Asymptotic Properties

All asymptotic theory is based on laws of large numbers on the one hand and central limit principles on the other. In this paper we will restrict ourselves to the estimation of ergodic scalar diffusions.

5.1 Ergodicity

Criteria for ergodicity can be given as follows: defining the scale measure

$$s(x,\theta) = \exp\,[\,-2\int_l^x \frac{\mu(y,\theta)}{\sigma^2(y,\theta)}dy]$$

for $b \in (l,r)$, we have the following conditions:

Condition For all θ in the parameter space, the following hold:

$$\int_l^b s(x,\theta)dx = \int_b^r s(x\theta)dx = \infty$$

and

$$A(\theta) = \int_l^r [s(x,\theta)\sigma^2(x,\theta)]^{-1}dx < \infty.$$

Under these conditions the process X_t is ergodic, has a stationary probability distribution with density

$$\mu_\theta(x) = [A(\theta)s(x,\theta)\sigma^2(x,\theta)]^{-1}$$

for $x \in (l,r)$ (Sorensen 2009).

5.2 Martingale Estimating Functions

Martingale estimating functions are functions of the type

$$G_\theta(\theta) = \sum_{t=0}^{T} g(X_t, X_{t-1}, \theta)$$

where $E[g(X_t, X_{t-1}, \theta)|X_{t-1}] = 0$. Since $E[\phi(X_{t+s}) - exp(\lambda s)\phi(X_t)|X_t] = 0$ Kessler and Sorensen (1999) have proposed using them to estimate eigenfunctions. In the general case they have considered functions of the form

$$H_T^k(\theta) = \sum_{t=0}^{T}\sum_{j=1}^{k}\beta_j(\theta)\phi_j(X_t, \theta) - exp(\lambda_j(\theta)\phi_j(X_{t-1}, \theta)$$

where the $\beta_j(\theta)$ are continuously differentiable functions of θ only and are chosen to minimize the variance of consistent and asymptotically normal estimators. Explicit expressions for β_j are given in Kessler and Sorensen (1999) and Sorensen (2009) for ergodic diffusions.

5.3 Convergence and Asymptotic Distribution

To show the convergence of the estimators and to find their asymptotic distributions,we will verify the conditions of theorems 2.2 and 2.3 of Sorensen (2009) as summarized below.

$$G_\theta(\theta) = \sum_{t=0}^{T} g(X_t, X_{t-1}, \theta)$$

is the objective function where $g(X_t, X_{t-1}, \theta)$ are martingale estimating functions. Let Q_θ denote the probability measure on the state space D of X_t with the density $p(x, y, \theta)$ of two consecutive observations.

$$Q_\theta(g_j(\theta)^2 = \int_{D^2} g_j(y, x, \theta)^2 p(x, y, \theta)dydx < \infty$$

where g_j are the components of g. The function $g(X_t, X_{t-1}\theta)$ is integrable with respect to Q_θ and that $Q_\theta(g(\theta_0) = 0$.

The function $g(x, y, \theta)$ is twice continuously differentiable, g and $\partial g_i(x, y, \theta)/\partial\theta_i$ $i = 1, ..., p$ are dominated by functions free from θ and integrable with respect to Q_θ.The p by p matrix $W = Q_{\theta_0}(\partial g_i(x, y, \theta)/\partial\theta)$ is of full rank. Then a consistent estimator exists with

$$\sqrt{T}(\hat{\theta}_T - \theta_0) \rightarrow N_p(0, W^{-1}VW^{T^{-1}}).$$

The matrices W and V can be estimated as follows:

$$W_n = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial_\theta g(X_t, X_{t-1}\hat{\theta})}{\partial\theta} \rightarrow W$$

and

$$V_n = \frac{1}{T}\sum_{t=1}^{T} g(X_t, X_{t-1}\hat{\theta})g(X_t, X_{t-1}, \hat{\theta})^T) \to V$$

where the convergence is with respect to the law of θ_0. All these results lead us to the following conclusion:

Theorem

The transition density and the eigenfunctions of an ergodic scalar stochastic differential equations

$$dX_t = \mu(X_t, \theta)dt + \sigma(X_t, \theta)dW_t$$

the solution of which is on a compact interval $[l, r]$ with two reflective barriers, can be estimated by the martingale estimation functions

$$\frac{1}{T}\sum_{t=0}^{T}[u(X_{t-1}, \gamma) - h(X_t)]$$

and

$$\frac{1}{T}\sum_{t=0}^{T}[\int_D \phi(y, \theta)p(y, X_{t-1}, \hat{\gamma})dy - exp(\lambda)\phi(X_t)]$$

where $\theta \in \theta$ a compact set. The estimators are convergent and asymptotically normal.

Proof

Taking

$$g(X_t, X_{t-1}, \theta) = u(X_{t-1}, \gamma) - h(X_t)$$

and then taking

$$g(X_t, X_{t-1}, \theta) = \int_D \phi(y, \theta)p(y, X_{t-1}, \hat{\gamma})dy - exp(\lambda)\phi(X_t)$$

theorems 2.2 and 2.3 of Sorensen (2009) will give us the desired results.

Q.E.D.

6 Conclusions

We have proposed a new way of estimating the transition density of a scalar stochastic differential equation which we have then used to estimate the eigenfunctions of the equations followed by the underlying process and the one followed by its eigenfunctions. The drift and the diffusion function can then be identified as indicated in section 3.

References

[1] Ait-Sahalia, Y. (2002). Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach. Econometrica, 70:223 - 262.

[2] Ait-Sahalia, Y. Mykland, P. (2003). The effects of random and discrete sampling when estimating continuous-time diffusions. Econometrica, 71:483-549.

[3] Ait-Sahalia, Y. Mykland, P. A. (2004). Estimators of diffusions with randomly spaced discrete observations: a general theory. Ann. Statist., 32:2186-2222.

[4] Chan, K. C.; Karolyi, G. A.; Longstaff, F. A. Sanders, A. B. (1992). An empirical comparison of alternative models of the short-term interest rate. Journal of Finance, 47:1209-1227.

[5] Cybenko, G. (1989). "Approximations by superpositions of sigmoidal functions", Mathematics of Control, Signals, and Systems, 2 (4): 303-314.

[6] Demoura, S. G. (1998). "The nonparametric estimation of the expected value operator" manuscript.

[7] Hall, P. Heyde, C. C. (1980). Martingale Limit Theory and Its Applications. Academic Press, New York.

[8] Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. Econometrica, 50:1029-1054.

[9] Hansen, L. P. Scheinkman, J. A. (1995). Back to the future: generating moment implications for continuous-time Markov processes. Econometrica, 63:767-804.

[10] Hansen, L. P. Scheinkman, J. A. Ait-Sahalia, Y. (2010) "Operator Methods for Continuous-Time Markov Processes" Handbook of Financial Econometrics Elsevier.

[11] Hornik, K. , M. Stinchombe, H. White (19890). "Multilayer Feedforward Networks are Universal Approximators" . Neural Networks, 2: 359-366.

[12] Hornik, K. , M. Stinchombe, H. White (1990) . "Universal Approximations of an Unknown Mapping and its Derivatives Using Multilayer Feedforward Networks". Neural Networks, 3: 551-560.

[13] Hornik, K. (1991), "Approximation Capabilities of Multilayer Feedforward Networks". Neural Networks, 4: 251-257.

[14] Gallant, A. , H. White (1992), "On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks". Neural Networks. 5: 129-138.

[15] Jacod, J. Sorensen, M. (2008). Asymptotic statistical theory for stochastic processes: a review.. Preprint, Department of Mathematical Sciences, University of Copenhagen. In preparation.

[16] Kessler, M. Sorensen, M. (1999). Estimating equations based on eigenfunctions for a discretely observed diffusion process. Bernoulli, 5:299-314.

[17] Sorensen, H. (2001). Discretely observed diffusions: Approximation of the continuoustime score function. Scand. J. Statist., 28:113-121.

[18] Sorensen, H. (2009). Parametric inference for discretely sampled stochastic dierential equations. In Andersen, T.G. Davis, R.A., Kreiss, J.-P. and Mikosch, T. (eds.): Handbook of Financial Time Series, Springer, Heidelberg, 2009, 531-553..