BANCO DE PORTUGAL

Economics Research Department

Simulated Likelihood Estimation of Non-Linear Diffusion Processes through Non-Parametric Procedure with an Application to the Portuguese Interest Rate

João Nicolau

WP 4-99 July 1999

Please address correspondence to the author at Instituto Superior de Economia e Gestão, e-mail: jnicolau@mail.telepac.pt.

Simulated Likelihood Estimation of Non-Linear Diffusion Processes through Non-Parametric Procedure with an Application to the Portuguese Interest Rate

João Nicolau*
Instituto Superior de Economia e Gestão

1999

Abstract

In this article we present a new model of the spot interest rate and a new method of estimation of nonlinear stochastic differential equations. We show how an integrated discrete time process in an econometric sense can be modelled by a continuous time ergodic process. We make an application to the Portuguese spot interest rate.

1 Introduction

In order to price derivatives securities, several stochastic differential equations (SDEs) have been proposed for the underlying traded asset(s) or non-traded factor(s). Usually, these SDEs are very simple (for instance, they are linear) and we can point out at least two reasons for this: first, simple SDEs can lead to explicit solutions to the prices of securities; and second, simple SDEs are easy to estimate, for instance, through maximum likelihood.

However, simple linear SDEs are commonly too crude to model the non-linearities of the financial time series data. It is important to stress some of the main features observed in financial time series data [see, for instance, Engle, Bollerslev e Nelson (1993)]. Let X_t be a financial time series data and $Y_t = X_t - X_{t-1}$ the first difference sequence of X. It is common to observe: (i) Y_t is strongly leptocurtic and, in certain cases, the unconditional variance is infinite, (ii) Y_t is, in general, uncorrelated, but not independent, (iii) the conditional variance of X_t and Y_t are not constant over time [in effect, «[...]large changes

*The author would like to thank Carlos Braumann and Nuno Cassola for providing help and guidance and to an anonymous referee for helpful comments. I am responsible for any remaining errors. This work was carried out at Bank of Portugal as part of a contract between CEMAPRE/ISEG and Bank of Portugal. Email: jnicolau@mail.telepac.pt

tend to be followed by large changes, of either sign, and small changes tend to be followed by small changes [...]» - Mandelbrot referred in Engle, Bollerslev e Nelson (1993)] (iv) X_t usually shows a very slow reversion to its mean or even do not have reversion at all (for instance the process is integrated of order one). These general features are well captured by some discrete-time models, for example by the GARCH model (applied to Y_t) but in a continuous-time framework it is hard to find a good general model [nevertheless, has recently been some promising work, see Aït-Sahalia (1996), Brandt and Santa-Clara (1999) among others].

On the other hand, non-linear SDEs are difficult to estimate and usually the maximum likelihood method it is not possible.

In this article we present a new spot interest rate model that takes into account some of features observed in the data. We present also a new method of estimation of SDEs that can be applied to any Markov processes (for instance, to dynamic discrete-time model). This model and this method are applied to the Portuguese spot interest rate.

2 The Model

When we admit that the data can be modelled by a diffusion model it is necessary to choose the drift $a(x,\theta)$ and diffusion the $b(x,\theta)$ coefficients (where θ represents a vector of parameters, such that $\theta \subset \mathbf{R}^k$). Most interest rate models specify a linear mean reverting drift, $a(x,\theta) = \alpha(\tau - x)$: so, the process moves towards τ at a speed $\alpha > 0$. Obviously, this not a logical assumption if X is an integrated process. In effect, in this case, we would expect a null drift. On the other hand there is evidence in some financial time series data, like the case of interest rate [see Aït-Sahalia (1996)] that the drift is null only in a certain region (for example, when the interest rate is low) and tends to be not null out of that region (for example, tends to be negative when the interest rate is high - which push the process towards the low interest rate region). See also Nicolau (1999). In this situation it is interesting to note that usually the Dicker-Fuller test does not reject the H0 of non-stationary.

In order to model the (logarithm of) spot interest rate process we propose the following model

$$dX_t = \alpha(\tau - X_t)\,dt + \exp\left\{\frac{\sigma}{2} + \frac{\beta}{2}(X_t - \mu)^2\right\} dW_t \qquad (1)$$

where, $t \in [0,T]$, $\alpha \geq 0$, $\sigma, \beta, \tau, \mu \in \mathbf{R}$ and W is a standard Wiener process. We will show that (1) with $\alpha = 0$ and $\beta > 0$ is a good model for the Portuguese interest rate process (and perhaps for others interest rates processes) in that it takes into account some of the non-linearities observed in the drift and in the volatility. However, for generality we assume $\alpha \geq 0$ and $\beta \neq 0$.

Proposition 1 *If $\alpha \geq 0$ the SDEs (1) has a global continuous solution for $t \in [0,\infty)$.*

Proof. See Appendix.

Proposition 2 *The solution of (1) admits a stationary density (SD) $\bar{p}(x)$ in the following cases:*

1. *$\alpha > 0$ and $\forall\beta$. The SD is*

$$\begin{aligned}\bar{p}(x) \quad &\propto \quad \exp\left\{\frac{\alpha}{\beta}e^{-\sigma-\beta(x-\mu)^2} - \sigma - \beta(x-\mu)^2\right. \\ &\left. + \frac{\alpha\sqrt{\pi}(\tau-\mu)}{e^{\sigma}}\frac{\operatorname{erf}\left(\sqrt{\beta}(x-\mu)\right)}{\sqrt{\beta}}\right\}\end{aligned}$$

where $\operatorname{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-u^2}du$.

2. *$\alpha = 0$ and $\beta > 0$. The SD is*

$$\bar{p}(x) = \frac{\exp\left\{-\beta(x-\mu)^2\right\}\sqrt{\beta}}{\sqrt{\pi}}$$

(in this case the SD is Gaussian with mean μ and variance $(2\beta)^{-1}$).

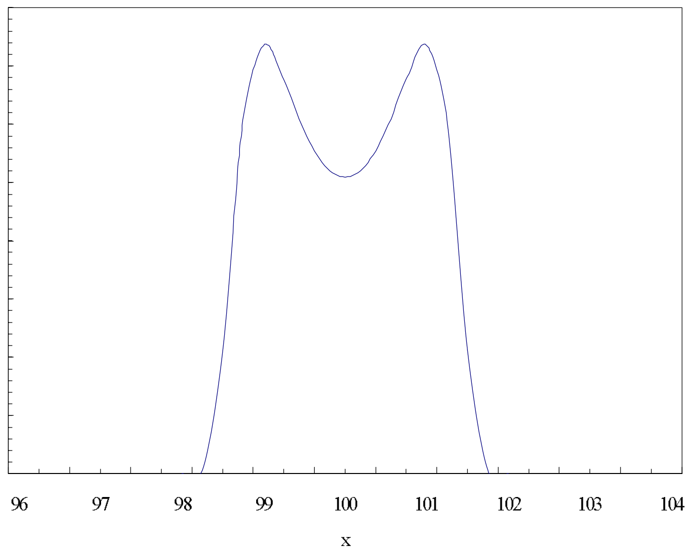
Proof. See Appendix.

There are several observations to make.

First, it is very surprising that the case $\alpha = 0$ and $\beta > 0$ admits a stationary density (SD) with stationary moments of any order. At first sight, with $\alpha = 0$ (i.e., with a null drift), the process should not have any attraction towards a stable point. However, the process tends to μ as $t \to \infty$. Moreover, $E[X_t | X_s] = X_s$ $(t \geq s)$ but the process is not integrated in the usual econometric sense because integrated processes diverge almost surely. Intuitively we can explain it as follows: when the process is near μ the instantaneous volatility, $b(x,\theta)$ is low and the process tends to remain near μ. If X drifts away from μ volatility increases. Now X is much more irregular so there is a positive probability that the process crosses μ again. It is the volatility that pushes the process towards a steady point. Technically, when the process is in its natural scale (see the following notation in the Appendix), i.e., $s(x) = 1$, the quantity $m(x)\varepsilon^2$ is of the order of the expected time the process spends in the interval $(x-\varepsilon, x+\varepsilon)$ given $X_0 = x$ before departure thereof [see Karlin e Taylor (1981), pp. 197-198; in effect $E[T_{x-\varepsilon,x+\varepsilon} | X_0 = x] = m(x)\varepsilon^2$ where $T_{a,b} = \min\{T_a, T_b\}$ and T_a is hitting time of a, so $T_{a,b}$ is the first time the process reaches either a or b]. It can be proved (see Appendix) that $m(x) = \exp\left\{-\beta(x-\mu)^2\right\}$, so $m(x)$ is maximum when $x = \mu$. That is, the process spends more time in the interval $(\mu-\varepsilon, \mu+\varepsilon)$ than in any other interval (with fix ε). Notice that according to previous preposition the unconditional mean of X under the hypothesis $\alpha = 0$ is μ.

Second, (with $\alpha \geq 0$ and $\beta > 0$) in this model «[...] large changes tend to be followed by large changes, of either sign, and small changes tend to be followed

Figure 1: Stationary Density ($\alpha = 1$, $\sigma = 1$, $\beta = -1$ and $\tau = 100$)



by small changes [...]». In effect, when X is near μ «[...] small changes tend to be followed by small changes [...]» as volatility is low; when X is far from μ «[...] large changes tend to be followed by large changes [...]» as volatility is high.

Third, the SD of the Ornstein-Uhlenbeck process, $dX_t = \alpha(\mu - X_t)\,dt + \sigma dW_t$ with $\sigma = \sqrt{\alpha/\beta}$ is equal to that of the $dX_t = \exp\left\{\frac{\sigma}{2} + \frac{\beta}{2}(X_t - \mu)^2\right\} dW_t$ process, so different dynamics can produce the same SD. It is clear that the SD could not identify the SDEs. As a consequence it seems prudent not to make inference based only on the distribution of the date.

Fourth, a very interesting case turns out to be $\alpha > 0$ and $\beta < 0$ (to simplify consider $\tau = \mu$). If $\sigma > \log(-\alpha/\beta)$ it is easy to show that the equation $dm(x)/dx = 0$ (note $\bar{p}(x) \propto m(x)$) has three real solutions (one of them is τ) so the stationary distribution is (bimodal). Why does this happen? The drift $\alpha(\tau - x)$, $\alpha > 0$, pushes the process towards τ, but near $\tau = \mu$ volatility is at its maximum. So at this point the process is repelled to the left or to the right of τ. To have intuition we plot (see figure 1) $m(x) = \exp\left\{\alpha \exp\left\{-\sigma - \beta(x-\tau)^2\right\}/\beta - \sigma - \beta(x-\tau)^2\right\}$ which gives the density $\bar{p}(x)$ for the following values $\alpha = 1$, $\sigma = 1$, $\beta = -1$ and $\tau = 100$.

If additionally we admit $\tau \neq \mu$ it can be proved that the SD is skewed. This model can also be useful to model an exchange rate process within a target zone where volatility is higher in the centre of the band and low near the bands.

3 A Method of Estimation of non-linear SDEs

The estimation of non-linear SDEs is generally difficult as the transition density (or the conditional density of X_t given $X_s = x$) and the conditional moments are usually not known. However, consistent methods have been proposed in those cases [see Duffie and Singleton (1993), Chan et al. (1992), Pedersen (1995), Hansen and Scheinkman (1995), Gouriéroux, Monfort and Renault (1995), Gallant and Long (1997), Aït-Sahalia (1997), Brandt and Santa-Clara (1999), among others]. The majority of these estimators are built into a stationarity framework and only a few of them are able to reach full efficiency.

Without assuming stationarity we are going to present a new method to estimate the parameters of a general non-linear diffusion process. Under certain conditions, the estimator is consistent and is as efficient as the maximum likelihood estimator.

Let $(X_{t_0}, X_{t_{1,...,}} X_{t_n})$ be the observed data at time $t_0, t_1, ..., t_n$. We suppose, for simplicity that $\Delta = t_i - t_{i-1}$ is constant. Let $p(\Delta, x, y; \theta) = \partial P\left[X_\Delta \leq y | X_0 = x\right] / \partial y$ be the true (but unknown) transition density

We propose to change the transition density, $p(\Delta, x, y; \theta)$ by an estimate $\hat{p}(\Delta, x, y; \theta)$ and solve the optimization problem

$$\max_\theta \frac{1}{n} \sum_{i=1}^{n} \log \hat{p}\left(\Delta, X_{t_{i-1}}, X_{t_i}; \theta\right) \tag{2}$$

in order to obtain a consistent estimator $\hat{\theta}_n$ of θ.

3.1 Transition Densities Estimation

We estimate $p\left(\Delta, X_{t_{i-1}}, X_{t_i}; \theta\right)$ non-parametrically by

$$\hat{p}\left(\Delta, X_{t_{i-1}}, X_{t_i}; \theta\right) = \frac{1}{n_s h} \sum_{j=1}^{n_s} K\left(\frac{X_{t_i} - Y_{t_i,j}(\theta)}{h}\right) \tag{3}$$

where, $K(\cdot)$ is a *kernel*, h is scalar (*bandwidth*) and $Y_{t_i,j}(\theta)$ is the j-th approximation of X at time t_i given the value of $X_{t_{i-1}}$.

Proposition 3 *Suppose that (i) the distribution of* Y *at time* Δ *given* x_0 *at time* $t_0 = 0$ *is equal to that of* X_Δ *given* x_0, *(ii)* $p(\Delta, x_0, x; \theta)$ *has a continuous second derivative with respect to* x, *(iii)* $\int K(u)\, du = 1$, $\int uK(u)\, du$, $\int u^2 K(u)\, du > 0$ *and (iv)* $n_s h \to \infty$, $h \to 0$. *Then*

$$\hat{p}(\Delta, x_0, x; \theta) \xrightarrow{q.m.} p(\Delta, x_0, x; \theta). \tag{4}$$

Proof. See Appendix.

To estimate $p\left(\Delta, X_{t_{i-1}}, X_{t_i}; \theta\right)$ we simulate n_s paths from t_{i-1} to t_i given $X_{t_{i-1}}$ and θ in order to obtain n_s approximations of X at time t_i given $X_{t_{i-1}}$.

So $(Y_{t_i,1}(\theta), Y_{t_i,2}(\theta) ..., Y_{t_i,n_s}(\theta))$ must give good information about the conditional mean, conditional variance and conditional distribution of X_{t_i}. Now we use the standard non-parametric tools to have an estimative of $p\left(\Delta, X_{t_{i-1}}, X_{t_i}; \theta\right)$.

There is a crucial point: $Y_{t_i,j}(\theta)$ must be simulated without any errors of approximations in order to assure that $Y_{t_i,j}(\theta)$ has the same conditional distribution as X_{t_i}. As in general we do not have the explicit solution of X_t it could be hard to assure that assumption. For instance, the Euler approximation could be a bad solution as, in this case, $Y_{t_i,j}(\theta) \sim N\left(a(x)\Delta, b^2(x)\Delta\right)$ given $X_{t_{i-1}} = x$. In order to have $Y_{t_i,j}(\theta)$ according to our assumptions we approximate X_Δ given $X_0 = x$ by $Y_\Delta^{(N)}$, last value of the following iterative procedure (cutting the j index):

$$Y_0^{(N)} = X_0 = x$$

$$Y_{k\frac{\Delta}{N}}^{(N)} = Y_{(k-1)\frac{\Delta}{N}}^{(N)} + a\left(Y_{(k-1)\frac{\Delta}{N}}^{(N)}\right)\frac{\Delta}{N} + b\left(Y_{(k-1)\frac{\Delta}{N}}^{(N)}\right)\sqrt{\frac{\Delta}{N}}\varepsilon_{k\frac{\Delta}{N}} \tag{5}$$

where $\varepsilon_{k\frac{\Delta}{N}} \sim N(0,1)$ and $k = 1, 2, ..., N$.

Let $p^{(N)}(\Delta, x, y; \theta)$ be the transition density of $Y^{(N)}$. We have the following results from Pederson (1995):

Proposition 4 *Assume for all $\theta \in \Theta$ that $a(x,\theta)$ and $b(x,\theta)$ are bounded with bounded derivatives of any order. Furthermore assume that there exists an $\varepsilon(\theta) > 0$ such that $b^2(x,\theta) - \varepsilon(\theta) \geq 0$ for all x. Then $p(\Delta, x, y; \theta)$ exists and*

$$p^{(N)}(\Delta, x, \cdot; \theta) \to p(\Delta, x, \cdot; \theta) \text{ in } L(\lambda) \text{ as } N \to \infty \tag{6}$$

(where λ is a Lebesgue measure).

Proof. See Theorem 3 in Pedersen (1995).

As Pedersen points out, these assumptions are very restrictive and it is possible to replace them by Lipschitz and growth like conditions. In effect only a few derivatives of a and b are really needed. On the other hand it can be shown, with simple regularity conditions, that [see Kloeden and Platen (1992)]

$$\lim_{N\to\infty} E\left[\left|X_\Delta - Y_\Delta^{(N)}\right|\right] = 0, \tag{7}$$

so we must have at time Δ, $Y^{(N)} \xrightarrow{d} X$ as $N \to \infty$.

For stochastic discrete-time models we can always simulate X_t given X_{t-1} without errors of approximation.

3.2 The Estimator

3.2.1 Algorithm

To obtain $\hat{\theta}_n$, the following optimization problem must be solved:

$$\max_\theta \log \hat{L}_n(\theta)$$

where

$$\log \hat{L}_n(\theta) = \frac{1}{n}\sum_{i}^{n} \log\left[\frac{1}{n_s h}\sum_{j=1}^{n_s} K\left(\frac{X_{t_i} - Y_{t_i,j}^{(N)}(\theta)}{h}\right)\right] \tag{8}$$

The algorithm is:

1. Simulate $\{\varepsilon_{i,j,\ell}; i = 1, ..., n; j = 1, ..., n_s; \ell = 1, ..., N\}$ where $\varepsilon_{i,j,\ell}$ is i.i.d. with gaussian distribution $N(0,1)$.
2. Let $k = 1$.
3. If $k = 1$ select a initial value for θ, (let $\theta^{(1)}$ be this value); if not fix $\theta^{(k)} = \theta^*$.
4. Given the initial condition $X_{t_0} = x$ and $\theta^{(k)}$, simulate n_s approximations for X_{t_1} using equation (5), that is $Y_{t_1,j}^{(N)}$ for $j = 1, ..., n_s$ (to each n_s approximation it is necessary take N random values $\varepsilon_{1,j,\ell}$, $\ell = 1, ..., N$). From $\left\{Y_{t_1,j}^{(N)}; j = 1, ..., n_s\right\}$ estimate $p\left(\Delta, X_{t_0}, X_{t_1}; \theta^{(k)}\right)$ using equation (3).
5. Repeat the last procedure until $p\left(\Delta, X_{t_{n-1}}, X_{t_n}; \theta^{(k)}\right)$ is estimated.
6. Calculate $\log \hat{L}_n\left(\theta^{(k)}\right) = \sum_{i=1}^{n} \log \hat{p}\left(\Delta, X_{t_{i-1}}, X_{t_i}; \theta^{(k)}\right)$.
7. Through a maximization procedure[1] select a new value θ^* such that $\log \hat{L}_n(\theta^*) \geq \log \hat{L}_n\left(\theta^{(k)}\right)$.
8. If $\left|\log \hat{L}_n(\theta^*) - \log \hat{L}_n\left(\theta^{(k)}\right)\right| < \varepsilon$ stop (ε is "small"). θ^* is the estimator. If not make $k = k + 1$ and return to step 3.

This procedure was implement with success in GAUSS[2]. The underlying principle in this method is the same as that of the maximum likelihood: given the observations, we select θ^*, function of the observations, that makes the sample more likely.

3.2.2 Consistency and Distribution

Let

$$\begin{aligned} Q_{n,n_s,N}(\theta) &= -\log \hat{L}_n(\theta) \\ &= -\frac{1}{n}\sum_{i}^{n} \log\left[\frac{1}{n_s h}\sum_{j=1}^{n_s} K\left(\frac{X_{t_i} - Y_{t_i,j}^{(N)}(\theta)}{h}\right)\right]. \end{aligned}$$

[1] We obtained good results from the following algorithms: Broyden-Fletcher-Goldfarb-Shanno, Newton-Raphson and BHHH.

[2] The GAUSS program can be given upon request.

We make the following assumptions: (A1) Θ is compact; (A2) the map $Q_{n,n_s,N}(\theta):\theta\to Q_{n,n_s,N}(\theta)$ is continuous; (A3) $Q_{n,n_s,N}(\theta)-Q_{n,n_s,N}(\theta_0)$ converge uniformly in θ to $\bar{Q}(\theta)-\bar{Q}(\theta_0)$ in q.m. as n, n_s and N tends to infinity; (A4) There exists a unique value θ_0 such that $\theta_0=\arg\min_{\theta\in\Theta}\bar{Q}(\theta)-\bar{Q}(\theta_0)$.

Under these assumptions there exist a sequence of extremum estimators $\widehat{\theta}_n$, solving the problem $\min\limits_{\theta\in\Theta} Q_{n,n_s,N}(\theta)$ that converge in q.m. to θ_0 (see Gouriéroux and Monfort (1995b), pp.387-388 with modifications).

To assure (A2) we impose the discretization scheme to be continuous in θ and the kernel $K(u)$ to be continuous in u.

Regarding (A3) we assume the conditions of proposition 3 which assure $\hat{p}\xrightarrow{m.q.}p$. Under these conditions

$$Q_{n,n_s,N}(\theta)=-\frac{1}{n}\sum_i^n\log\left[\frac{1}{n_s h}\sum_{j=1}^{n_s}K\left(\frac{X_{t_i}-Y_{t_i,j}^{(N)}(\theta)}{h}\right)\right]$$

converge in q.m. to

$$-\frac{1}{n}\sum_i^n\log p\left(\Delta,X_{t_{i-1}},X_{t_i};\theta\right)$$

as $N\to\infty$, $n_s h\to\infty$ and $h\to 0$ (according to the algorithm, it is legitimate to impose first, $N\to\infty$, $n_s h\to\infty$ and $h\to 0$ and finally $n\to\infty$). We assume, as an additional assumption that $-n^{-1}\sum_i^n\log p\left(\Delta,X_{t_{i-1}},X_{t_i};\theta\right)$ converge in θ uniformly to $E_{P^o}\left[\log p(\Delta,x,y;\theta)\right]$ in q.m. sense.

Finally we prove (A4). In effect, under (A3)

$$\begin{aligned}Q_{n,n_s,N}(\theta)-Q_{n,n_s,N}(\theta_0) &= -\frac{1}{n}\sum_i^n\log\left[\frac{1}{n_s h}\sum_{j=1}^{n_s}K\left(\frac{X_{t_i}-Y_{t_i,j}^{(N)}(\theta)}{h}\right)\right]\\ &\quad+\frac{1}{n}\sum_i^n\log\left[\frac{1}{n_s h}\sum_{j=1}^{n_s}K\left(\frac{X_{t_i}-Y_{t_i,j}^{(N)}(\theta_0)}{h}\right)\right].\end{aligned}$$

converge in q.m. to the Kullback information

$$K(\theta,\theta_0)=E_{P^o}\left[\log\frac{p(\Delta,x_0,x;\theta_0)}{p(\Delta,x_0,x;\theta)}\right]$$

as $N\to\infty$, $n_s h\to\infty$, $h\to 0$ and $n\to\infty$. As $K(\theta,\theta_0)\geq 0$ [see Gouriéroux and Monfort (1995a)] and $K(\theta,\theta_0)=0$ if $\theta=\theta_0$ the solution of the limit problem is, therefore, $\theta_0=\arg\min_\theta K(\theta,\theta_0)$.

Another possibility of proof can be based on the assumption that the maximum likelihood (ML) estimator exists. With the additional assumptions implied in $\hat{p}\xrightarrow{m.q.}p$ the optimization problem $\min_\theta Q_{n,n_s,N}(\theta)$ is equivalent, as $n_s h\to\infty$, $h\to 0$, to the ML problem $-\max Q_{n,\infty,\infty}(\theta)$ where $-Q_{n,\infty,\infty}(\theta)$ is the true log-likelihood function. So, the solution of $\min_\theta Q_{n,\infty,\infty}(\theta)$ leads to the maximum likelihood.

Suppose now that there exists the ML estimator such that

$$\sqrt{n}\left(\tilde{\theta}_n - \theta\right) \xrightarrow{d} N\left(0, I\left(\theta_0\right)^{-1}\right)$$

where $I(\theta_0)$ is the Fisher information matrix. Gouriéroux and Monfort (1991) have proved, assuming $\sqrt{n}/n_s \to 0$, that the estimator based on the simulation of the log-likelihood function has the same asymptotical distribution than that of the maximum likelihood. So we can estimate $I(\theta_0)$, for instance, by

$$\frac{1}{n}\sum_{i}^{n}\frac{\partial \log \hat{p}\left(X_{t_i} | X_{t_{i-1}}; \hat{\theta}_n\right)}{\partial \theta}\frac{\partial \log \hat{p}\left(X_{t_i} | X_{t_{i-1}}; \hat{\theta}_n\right)}{\partial \theta^T}.$$

As a final point it is important to observe that the consistency and efficiency of this estimator depends on how good is the estimate $\hat{p}$ of p, that is, depends of N and n_s and these parameters are under the researcher's control.

4 An Application to the Portuguese Spot Interest Rate

We studied the Portuguese spot interest rate (three months deposit - see figure 2).

Two adjacent outliers were dropped from the sample as they impair the estimation (we admit that they do not belong to the data generating process). We studied the logarithm of interest rate[3]. We present some statistics for $X_t = \log(r_t)$ where r_t is the Portuguese interest rate:

Table 1: Statistics

Variable	mean	variance	kurtosis	skewness	ADF1	ADF2	ADF3
X_t	2.05	.1159	1.82	-.24	-1.950	-.352	-3.239
$X_t - X_{t-1}$	-.0008	.00033	35.53	2.52	-29.24	-29.35	-29.35

ADF1: Augmented Dickey-Fuller Stat. for random walk without drift (c.v.=-1.957)
ADF2: ADF Stat. for random walk with drift (c.v.=-2.81)
ADF3: ADF Stat. for random walk with drift and trend (c.v.=-3.43)

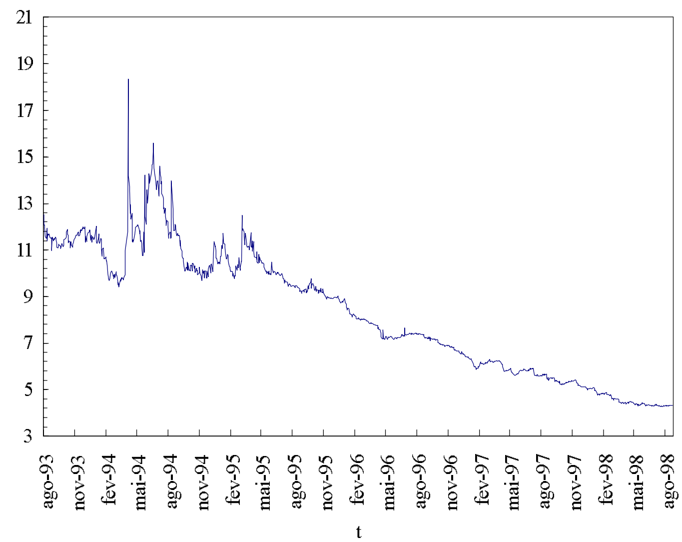
We can not reject the hypothesis that X is non-stationary. $X_t - X_{t-1}$ is stationary and displays high kurtosis.

Firstly, to have an idea about the behaviour of the drift and of the diffusion coefficients we estimate them non-parametrically by, respectively

$$M_n(x) = \frac{\sum_{i=1}^{n}\mathcal{I}_{\left\{\left|X_{t_{i-1}}-x\right|<h\right\}}\left(X_{t_i} - X_{t_{i-1}}\right)/\Delta}{\sum_{i=1}^{n}\mathcal{I}_{\left\{\left|X_{t_{i-1}}-x\right|<h\right\}}}$$

[3] This transformation is necessary as the solution of (1) could assume negative values.

Figure 2: Portuguese Interest Rate (three months deposit)



and

$$V_n(x) = \frac{\sum_{i=1}^{n} \mathcal{I}_{\left\{\left|X_{t_{i-1}} - x\right| < h\right\}} \left(X_{t_i} - X_{t_{i-1}}\right)^2 / \Delta}{\sum_{i=1}^{n} \mathcal{I}_{\left\{\left|X_{t_{i-1}} - x\right| < h\right\}}}$$

where $\mathcal{I}_{\left\{\left|X_{t_{i-1}} - x\right| < h\right\}} = 1$ if $\left|X_{t_{i-1}} - x\right| < h$ and zero otherwise[4]. $M(x)$ and $V(x)$ are, respectively estimates for $a(x)$ and $b^2(x)$. Florens-Zmirou (1993) has proved (with some regularity conditions) that $V_n(x) \xrightarrow{p} b^2(x)$ as $n \to \infty$, $h \to 0$, $nh \to \infty$ and $nh^4 \to 0$ and $\sqrt{N_x/2}\left(V_n(x)/b^2(x) - 1\right) \xrightarrow{d} N(0,1)$ where $N_x = \sum_{i=1}^{n} \mathcal{I}_{\left\{\left|X_{t_i} - x\right| < h\right\}}$. It can be proved that $M(x) \xrightarrow{p} a(x)$.

In figure 3 we observe that the drift is approximately null until the value 2.4 ($e^{2.4} \approx 11\%$ interest rate) and then decreases strongly. According to these estimates, it is not clear what is the equilibrium value. There are two points where the drift is null: the values 2.2 and 2.4 but it is possible that the equilibrium (if it exists) is at a lower level as, below 2.1 the drift is slightly negative. Nevertheless, when the process is below the value 2.4 it seems to behave like a random walk, but it displays reversion to lower values when the interest rate increases to high values. In figure 4 we observe a sharp increase in volatility

[4] To achieve more efficiency all estimates satisfy the conditions $\sum_{i=1}^{n} \mathcal{I}_{\left\{\left|X_{t_{i-1}} - x\right| < h\right\}} \geq 20$. It should be possible to consider other kernels, like the Gaussian. But we follow Florens-Zmirou (1993).

Figure 3: Non Parametric Drift Estimation

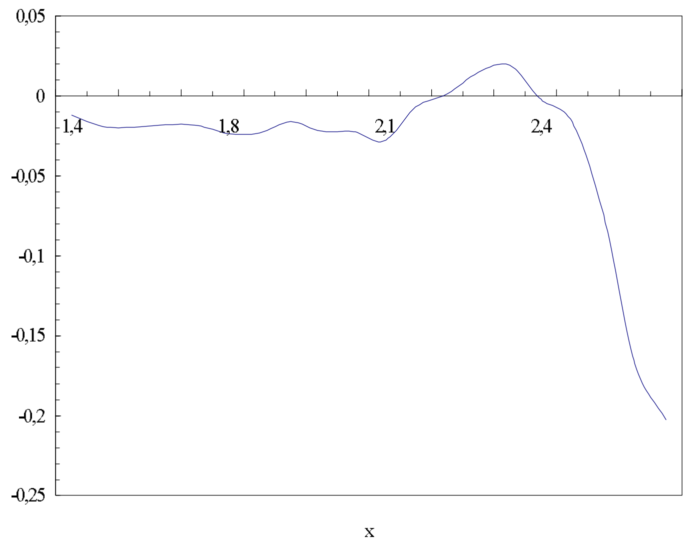
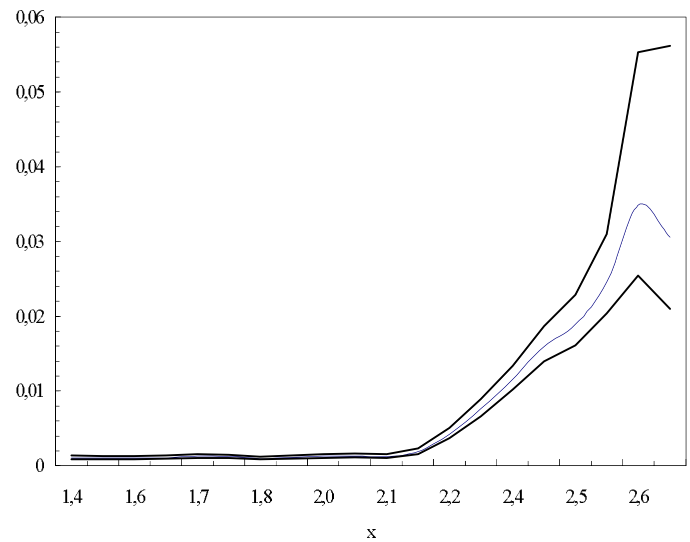


Figure 4: Non Parametric Diffusion Estimation and Conf. Inter.



above the value 2.15 ($e^{2.15} \approx 8.58\%$ interest rate). It is possible to admit a quadratic or exponential form to the diffusion coefficient but not linear one.

We perform the analysis with $t = 1$ meaning one month, so the daily interval is (approximately) 1/30. The interval between Monday and Friday (of the final week) is 3/30 (holidays are also accommodated).

We first tried the general model (1) but the parameter α was not significant. So we estimated (1) imposing the restriction $\alpha = 0$ (i.e., a model with null drift).

In the following table we present the results of the Euler approximation based estimation (i.e., the maximum likelihood applied to the Euler discrete equation)

Table 2[5]: Euler Approximation Based Estimation

Parameters	Estimates	t-ratio
σ	-6.72	-50.7
β	4.77	3.75
μ	1.67	21.6

Applying the method of estimation we described above with $N = 10$ and $n_s = 200$ we obtained

Table 3[6]: Consistent Method

Parameters	Estimates	t-ratio
σ	-6.60	-92.8
β	3.54	19.91
μ	1.52	44.27

There are some differences between the two results. The most important regard to the β and standard errors estimates.

According to the model, the estimate of μ gives the long-term mean of the process, so $e^{1.52} \approx 4.57$ is the long-term value of the interest rate with minimum volatility. Therefore, an integrated discrete-time series process, according to the Dickey-Fuller test, can be modelled by a continuous-time stationary process with a null drift.

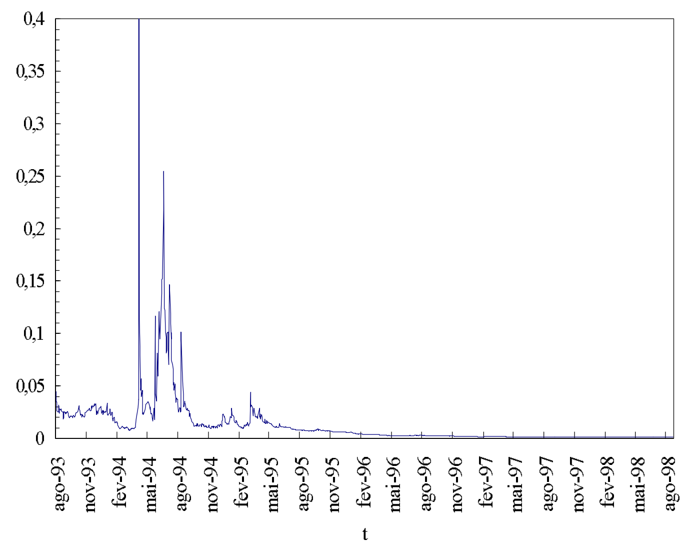
In figure 5 we present an approximation to the conditional variance[7], $\Delta b^2 \left(X_{t_{i-1}} \right)$, based on estimated parameters (table 3).

It remains to be seen if the model we have proposed can capture the main features of the data. To analyze the specification of the fitted model we perform the following bootstrap study:

1. simulate the process (1) with the estimated parameters (table 3);
2. obtain the non-parametric drift and diffusion estimates;

[5] The covariance of the parameters are computed by quasi maximum likelihood method.
[6] The covariance of the parameters are computed by quasi maximum likelihood method.
[7] It can be proved that $Var\left[X_{t_i} | X_{t_{i-1}} \right] = \Delta b^2 \left(X_{t_{i-1}} \right) + O\left(\Delta^2 \right)$.

Figure 5: Conditional Variance of log Interest Rate



3. repeat the last two steps 1000 times;
4. build a confidence interval from the empirical distribution of the drift and diffusion estimates.

The idea is to investigate if the fitted model is able to generate the drift and diffusion estimates obtained from the data, that is, as we had performed in figures 3 and 4. Figures 6 and 7 allow us to conclude that the fitted model is well specified.

It is interesting to observe that the underlying non-parametric drift of the model (1) with $a(x, \theta) = 0$ has a strong tendency to be negative above the value 2.4 (it is necessary to distinguish between the underlying non-parametric drift and the true drift which is null over the state space). That is, despite $a(x, \theta) = 0$ the process has a reversion to lower values when the process assumes high values (see the discussion at beginning).

5 Conclusion

We have presented a new model and a new method of estimation which we applied to Portuguese interest rate. One important conclusion is that an integrated process in the usual econometric sense can be a stationary continuous-time ergodic process with a stationary distribution.

Figure 6: Non Paramet. Drift Estimation and Conf. Inter. Obtained from Bootstrap Application

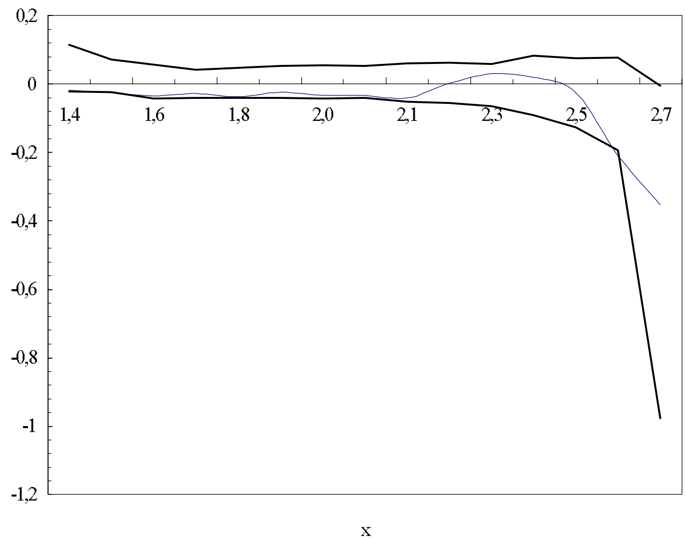
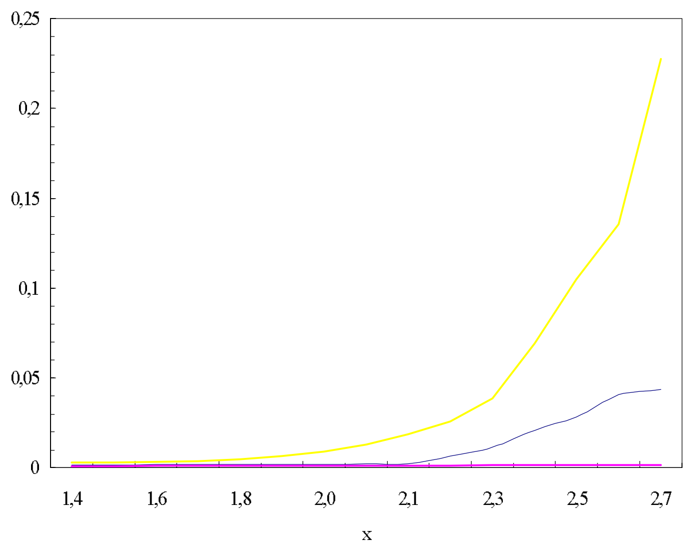


Figure 7: Non Paramet. Diffusion Estimation and Conf. Inter. Obtained from Bootstrap Application



We have concluded that the logarithm of Portuguese interest rate behaves like a random walk when the interest rate is low. However, when it is high, Portuguese interest rate display a reversion to lower values. We observe low volatility within a certain interval of the state space. However, volatility increases quickly when the process, is out of that region. This non-linearity can be captured by an exponential volatility model with a null drift. Our specification analysis does not reject the proposed model.

One of the basic ideas of this article, namely, that the volatility can push the process towards a steady point even if the drift coefficient is null, can be explored by other models. We will address this question in the future.

References

Aït-Sahalia Y., (1996), "Testing Continuous-Time Models of the Spot Interest Rate", *The Review of Financial Studies*, Vol. 9, N° 2, pp.385-426.

Aït-Sahalia Y., (1997), "Maximum Likelihood Estimation of Discretely Sampled Diffusions: a Closed-Form Approach", unpublished manuscript.

Arnold L., (1974), *Stochastic Differential Equations: Theory And Application*, John Wiley & Sons.

Brandt M.W., Santa-Clara P., (1999), "Simulated Likelihood Estimation of Multivariate Diffusions with an Application to Interest Rates and Exchange Rates with Stochastic Volatility", unpublished manuscript.

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", *The Journal of Finance*, Vol.XLVII, n3.

Duffie D., Singleton K., (1993), "Simulated Moments Estimation of a Markov Models of a Asset Prices", *Econometrica*, 50, 987-1007.

Engle R. F., Bollerslev T., Nelson D.B. (1993), "ARCH Models", *Handbook of Econometrics*, Vol.4, Chap. 49.

Florens-Zmirou D., (1993), "On Estimating the Diffusion Coefficient from Discrete Observation", *Journal of Applied Probability*, 30, pp. 790-804 .

Gallant A.R., Long J.R., (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum chi-squared", *Biometrica*, 84, 1.

Gouriéroux C., Monfort A., (1991), "Simulation Based Inference In Models With Heterogeneity", Annales D'conomie et de Statistique, n20/21.

Gouriéroux C., Monfort A., (1995a), *Statistics and Econometric Models*, Vol. I, Cambridge University Press.

Gouriéroux C., Monfort A., (1995b), *Statistics and Econometric Models*, Vol. II, Cambridge University Press.

Gouriéroux C., Monfort A., Renault E., (1995), "Indirect Inference", in Dijk H.K.V., Monfort A., Brown B.W. (ed.), *Econometric Inference Using Simulation Techniques*, Jonh Wiley and Sons.

Hansen L.P., Scheinkman J.A., (1995), "Back to the Future: Generating Moment Implications for Continous-Time Markov Processes", *Econometrica*, 63, n^{o}4.

Ikeda N., Watanabe S. (1981), *Stochastic Differencial Equations and Diffusion Processes*, North-Holland Mathematical Library.

Karlin S., Taylor H.M., (1981), *A Second Course in Stochastic Processes*, Academic Press.

Kloeden P.E., Platen E., (1992), *Numerical Solution of Stochastic Differencial Equations*, Springer-Verlag.

Nicolau J., (1999), "Modelling the DEM/USD Exchange Rate as an Ergodic Stationarity Continuous Time Process", Working Paper n^{o}2-99, CEMAPRE - Instituto Superior de Economia e Gestão.

Pederson A. R., (1995), "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equation based on Discrete Observations", *Scandinavian Journal of Statistics*, 22, 55-71.

Skorokhod A.V, (1989), *Asymptotic Methods in the Theory of Stochastic Differential Equation*, Translation of Mathematical Monographs Vol. 78, American Mathematical Society.

Appendix

Let $dX_t = a(X_t)\,dt + b(X_t)\,dW_t,\ t > t_0$, be a diffusion process, $s(z) = \exp\left\{-\int_{z_0}^{z} 2a(u)/b^2(u)\,du\right\}$ the scale density function, $m(u) = \left(b^2(u)\,s(u)\right)^{-1}$ the speed density and $I = (l, r)$ the state of space of X process (see Karlin and Taylor (1981)]. Let $S(l, x] = \lim_{a\to l}\int_a^x s(u)\,du$ and $S[x, r) = \lim_{b\to r}\int_x^b s(u)\,du$. According Arnold (1974), p.114, if the infinitesimal coefficient a and b have continuous derivatives with respect to x, then there exists a unique continuous process defined until the random moment expulsion η in the interval $t_0 < \eta \leq \infty$. Ikeda and Watanable (1981), pp. 362-363 prove that if $S(l, x] = S[x, r) = \infty$ then

$$P\left[\eta = \infty \middle| X_0 = x\right] = P\left[\overline{\lim_{t\to\infty}}\, X_t = r \middle| X_0 = x\right] = P\left[\underline{\lim_{t\to\infty}}\, X_t = l \middle| X_0 = x\right] = 1.$$

Now, it is known that if $S(l, x] = S[x, r) = \infty$ and $\int_l^r m(x)\,dx < \infty$ then X is ergodic and the invariant distribution P^0 has density $\bar{p}(x) = m(x) / \int_l^r m(u)\,du$ with respect with Lebesgue measure [see Skorokhod (1989), pp. 46-48].

Proof of Proposition 1

From (1) we see that a and b have continuous derivatives with respect to x. So, we must prove that $S(l,x] = S[x,r) = \infty$ where $l = -\infty$ and $r = \infty$. We have

$$\begin{aligned} s(x) &= \exp\{\phi_1(x) + \phi_2(x)\} \\ &= \exp\left\{-\frac{\alpha}{\beta}e^{-\sigma-\beta(x-\mu)^2} + \frac{\alpha\sqrt{\pi}(\mu-\tau)}{e^{\sigma}}\frac{\operatorname{erf}\left(\sqrt{\beta}(x-\mu)\right)}{\sqrt{\beta}}\right\} \end{aligned}$$

where $\operatorname{erf}(x) = 2/\sqrt{\pi}\int_0^x e^{-u^2}du$ (notice: $-1 \le \operatorname{erf}(x) \le 1$ for all x). If $\beta > 0$, the function $\phi_1(x) + \phi_2(x)$ is limited for all x, thus $\lim_{x\to\pm\infty} s(x) > 0$ e $S(l,a] = S[b,r) = \infty$. If $\beta < 0$, the function $\phi_1(x)$ tends to infinity as $x \to \pm\infty$. In what concerns $\phi_2(x)$ it is necessary observe that ($i = \sqrt{-1}$)

$$\operatorname{erf}(ix) = \frac{2}{\sqrt{\pi}}\int_0^{ix} e^{-u^2}du = \frac{2}{\sqrt{\pi}}i\int_0^x e^{t^2}dt$$

(making the change $u = it$). So

$$\begin{aligned} \phi_2(x) &= \frac{\alpha\sqrt{\pi}(\mu-\tau)}{e^{\sigma}}\frac{\operatorname{erf}\left(i\sqrt{|\beta|}(x-\mu)\right)}{i\sqrt{|\beta|}} \\ &= \frac{\alpha\sqrt{\pi}(\mu-\tau)}{e^{\sigma}}\frac{\operatorname{erfi}\left(\sqrt{|\beta|}(x-\mu)\right)}{\sqrt{|\beta|}} \end{aligned}$$

where $\operatorname{erfi}(x) = 2/\sqrt{\pi}\int_0^x e^{t^2}dt$. The function $\phi_2(x)$ tends to $-\infty$ when $\mu < \tau$ and $x \to \infty$ or when $\mu > \tau$ and $x \to -\infty$. However, $\phi_1(x) + \phi_2(x)$ tends to ∞ as $x \to \pm\infty$ (notice that $d\left(e^{x^2} - \operatorname{erfi}(x)\right)/dx = 2e^{x^2}\left(-\pi^{-1/2} + x\right)$; thus, the derivative of $\phi_1(x) + \phi_2(x)$ is positive when x tends to ∞ and negative when x tends to $-\infty$). In both cases, we have $\lim_{x\to\pm\infty} s(x) > 0$ so $S(l,a] = S[b,r) = \infty$ (where a and b are any two points of I). ■

Proof of Proposition 2

We must prove that $\int_{\mathbf{R}} m(x)\,dx < \infty$. First, consider the case $\alpha > 0$. It can be shown that $m(x)$ is

$$m(x) = \exp\left\{\frac{\alpha}{\beta}e^{-\sigma-\beta(x-\mu)^2} - \sigma - \beta(x-\mu)^2 + \frac{\alpha\sqrt{\pi}(\tau-\mu)}{e^{\sigma}}\frac{\operatorname{erf}\left(\sqrt{\beta}(x-\mu)\right)}{\sqrt{\beta}}\right\}.$$

For $\beta > 0$, the quantity $\frac{\alpha}{\beta}\exp\left\{-\sigma - \beta(x-\mu)^2\right\}$ is approximately zero (when $|x|$ is moderately high); on the other hand, $e^{-\sigma}\alpha\sqrt{\pi}(\tau-\mu)\beta^{-1/2}\operatorname{erf}\left(\sqrt{\beta}(x-\mu)\right)$ is bounded. Thus, it is possible to find a function $f(x) = \exp\left\{k_1 - k_2x^2\right\}$ such that $\exists k_1, k_2 > 0 : f(x) \ge m(x)$. The convergence of $\int_{\mathbf{R}} f(x)\,dx$ imply the convergence of $\int_{\mathbf{R}} m(x)\,dx$. For $\beta < 0$,the $m(x)$ function can be written

$$m(x) = \exp\left\{\frac{\alpha}{\beta}e^{-\sigma-\beta(x-\mu)^2} - \sigma - \beta(x-\mu)^2 + \frac{\alpha\sqrt{\pi}(\tau-\mu)}{e^{\sigma}}\frac{\operatorname{erfi}\left(\sqrt{|\beta|}(x-\mu)\right)}{\sqrt{|\beta|}}\right\}$$

(notice: $\lim_{x\to-\infty} \text{erfi}(x) = \infty$ and $\lim_{x\to-\infty} \text{erfi}(x) = -\infty$). Consider, without any loss of generality the case $\beta = -1$, $\alpha = 1$, $\sigma = 1$, $\tau = 1$ and $\mu = 0$. Thus $m(x) = \exp\{\gamma(x)\} = \exp\left\{-e^{x^2} + x^2 + \text{erfi}(x)\right\}$.The $\gamma(x)$ function tends quickly to $-\infty$ (in effect, $\gamma'(x) = 2e^{x^2}(1-x) + 2x$ is positive when $x \to \infty$ and negative when $x \to -\infty$). Thus, it is possible find a function $f(x) = \exp\{k_1 - k_2 x^2\}$ such that: $\exists k_1, k_2 > 0 : f(x) \geq m(x)$. The convergence of $\int_{\mathbf{R}} f(x)\,dx$ imply the convergence of $\int_{\mathbf{R}} m(x)\,dx$. For $\beta = 0$, X is the Ornstein-Uhlenbeck process which is stationary. In the case, $\alpha = 0$, we have $m(x) = e^{-\sigma-\beta(x-\mu)^2}$. It is easily to see that $\int_{\mathbf{R}} m(x)\,dx < \infty$ if $\beta > 0$. ■

Proof of Proposition 3. For simplicity we do not write θ. To show the dependence of $\hat{p}$ from Y simulation process we write $\hat{p}\left(\Delta, X_{t_{i-1}}, X_{t_i}; Y_{t_i}\right)$ instead of $\hat{p}\left(\Delta, X_{t_{i-1}}, X_{t_i}\right)$. We have,

$$E_y\left[\hat{p}(\Delta, x_0, x; y) | x_0\right] = \int \hat{p}(\Delta, x_0, x; y)\, g(\Delta, x_0, y)\, dy.$$

where $g(\Delta, x_0, \cdot)$ is the density of Y given x_0. By hypothesis (i) $g(\Delta, x_0, y) = p(\Delta, x_0, y)$. Thus

$$\begin{aligned} E_y\left[\hat{p}(\Delta, x_0, x; y) | x_0\right] &= \int \hat{p}(\Delta, x_0, x; y)\, p(\Delta, x_0, y)\, dy \\ &= \frac{1}{n_s}\sum_{j=1}^{n_s} \int \frac{1}{h} K\left(\frac{x-y}{h}\right) p(\Delta, x_0, y)\, dy \\ &= \frac{1}{n_s}\sum_{j=1}^{n_s} \int K(u)\, p(\Delta, x_0, x - hu)\, du \end{aligned}$$

(making the variable change $y = x - hu$). By Taylor's formula

$$\begin{aligned} p(\Delta, x_0, x - hu) &= p(\Delta, x_0, x) - hup'(\Delta, x_0, x) + \frac{1}{2}h^2u^2p''(\Delta, x_0, x) \\ &\quad + O\left(h^3\right) \end{aligned}$$

so

$$\begin{aligned} E_y\left[\hat{p}(\Delta, x_0, x; y) | x_0\right] &= \frac{1}{n_s}\sum_{j=1}^{n_s} \int K(u)\, p(\Delta, x_0, x - hu)\, du \\ &= \frac{1}{n_s}\sum_{j=1}^{n_s} \left\{ p(\Delta, x_0, x) \int K(u)\, du \right. \\ &\quad - hp'(\Delta, x_0, x) \int uK(u)\, du \\ &\quad \left. + \frac{1}{2}h^2 p''(\Delta, x_0, x) \int u^2 K(u)\, du + \ldots \right\} \\ &= p(\Delta, x_0, x) \\ &\quad + \frac{1}{2}h^2 p''(\Delta, x_0, x) \int u^2 K(u)\, du + O\left(h^4\right). \end{aligned}$$

It can be shown, by the same reasons (notice that $\{Y_{t_i,j}, j = 1, 2, ..., n_s\}$ is a sequence of i.i.d. random variables given x_0), that

$$Var_y\left[\hat{p}\left(\Delta, x_0, x; y\right)\middle| x_0\right] = \frac{\int K^2\left(u\right) du}{n_s h} + O\left(n^{-1}\right)$$

It follows, by hypothesis (iv) that $E_y\left[\hat{p}\left(\Delta, x_0, x; y\right)\middle| x_0\right] \rightarrow p\left(\Delta, x_0, x\right)$ and $Var_y\left[\hat{p}\left(\Delta, x_0, x; y\right)\middle| x_0\right] \rightarrow 0.$ ■