Bull Braz Math Soc, New Series 33(3), 307-318
© 2002, Sociedade Brasileira de Matemática

Ergodicity, geometric ergodicity and mixing conditions for nonparametric ARMA processes*

Graciela Boente[1] and Ricardo Fraiman

Abstract. In this paper, we introduce nonparametric ARMA models which provide an alternative to nonparametric autoregressive models, when there is a large dependence to the past observations. Conditions for ergodicity and geometric ergodicity are given when both the nonparametric autoregressive part and the moving average structure depend only one step behind. Also, a Fisher–consistent procedure is provided and its performance is studied through a simulated example.

Keywords: ARMA models, geometric ergodicity, Fisher–consistency, kernel estimates.

Mathematical subject classification: 62G08, 37A25.

1 Introduction

Autoregressive models with moving average errors (ARMA models) have been extensively used in applications when dealing with time series data. They correspond to linear autoregressive models where the errors are described by a moving average process. More precisely, an ARMA (p, q) model, is a stationary process $\{X_t : t \geq 1\}$ verifying

$$X_t = \sum_{j=1}^{p} \phi_j X_{t-j} + \varepsilon_t \ , \tag{1}$$

where $\varepsilon_t = u_t - \sum_{j=1}^{q} \theta_j u_{t-j}$ with u_t i.i.d. random variables and u_t independent of $\{X_{t-j},\ j \geq 1\}$, $E|u_t| < \infty$.

Received 9 April 2002.

*This research was partially supported by Grant PICT # 03-00000-006277 from ANPCYT at Buenos Aires, Argentina.

[1]The research of Graciela Boente was partially supported by a Guggenheim fellowship.

It is well known that, quite often, ARMA models have several advantages with respect to autoregressive models, when there is a large dependence to the past observations. As pointed by Granger and Newbold (1986) “If some economic variable is in equilibrium but is moved from the equilibrium position by a series of buffering effects from unpredictable events either from within the economy, such as strikes, or from outside, such as periods of exceptional weather, and the system is such that the effects of such events are not inmediately assimilated, then a moving average model will arise”. In such situations, ARMA models allow to fit several real data sets by using less parameters by introducing the moving average structure for the errors ε_t. This is related to the “principle of parsimony” which suggests that parameters should be introduced only when they are needed.

On the other hand, the assumption of a linear autoregression function is quite restrictive. As pointed by Bosq (1996) a nonparametric predictor is “in general more efficient and more flexible than the predictor based on Box and Jenkins method and nearly equivalent if the underlying model is truly linear” (see also Carbon and Delecroix (1993) for a comparative study on 17 series).

However, the nonparametric autoregression model $X_t = m\,(\mathbf{X}_t) + u_t$, where $\mathbf{X}_t = (X_{t-1}, \ldots, X_{t-r})$, faces the problem known as the “curse of dimensionality”. In order to solve the problem of empty neighborhoods, an approach can be to introduce moving average errors which reduce the dependence to the past in $\mathbf{X}_t$ obtaining thus a smaller dimension r .

Putting things together, following a semiparametric approach, we introduce nonparametric ARMA models which allow the autoregressive part of the model to be nonparametric, while the moving average part remains linear.

By analogy to (1), one can consider a stationary process $\{X_t : t \geq 1\}$ verifying

$$X_t = g\left(X_{t-1}, \ldots, X_{t-p}\right) + \varepsilon_t \ , \tag{2}$$

where $\varepsilon_t = u_t - \sum_{j=1}^{q} \theta_j u_{t-j}$ with u_t i.i.d. random variables and u_t independent of $\{X_{t-j}, \, j \geq 1\}$, $E|u_t| < \infty$. From now on, we will refer to a stochastic process verifying (2) as a nonparametric ARMA (p, q) model and it will be denoted by NARMA (p, q) model.

In particular, the NARMA (1,1) is a stationary process $\{X_t : t \geq 1\}$ verifying

$$X_t = g(X_{t-1}) + \varepsilon_t \ , \tag{3}$$

with $\varepsilon_t = u_t - \theta_0 u_{t-1}$; u_t independent and identically distributed and u_t independent of $\{X_{t-j}, \ j \geq 1\}$, $E|u_t| < \infty$.

In this paper, we give conditions for ergodicity and geometric ergodicity of the NARMA (1,1) model. From this last statement, under Harris recurrence and aperiodicity of the chain, it follows that the process is also a geometric α-mixing process. As is well known, mixing conditions have shown to be useful to derive asymptotic properties of kernel estimates for nonparametric autoregression models (see for instance, Bosq (1996) and the references therein).

In Section 2 we prove under mild conditions the ergodicity of NARMA (1,1) processes, while in Section 3 we give conditions for geometric ergodicity and mixing properties of NARMA (1,1) processes. Finally, in Section 4 we propose Fisher–consistent estimates of the autoregression function g and the moving average parameter θ, and an iterative procedure to calculate them. We also illustrate its behaviour through two simulated examples.

We now introduce some notation. Let f be the density of u_t; and $f_{u_1|X_1=x}$ the density of $u_t \mid X_t = x$ with respect to the Lebesgue measure λ. Then, $\varepsilon_t \mid X_{t-1} = x$ has a density f_x given by

$$f_x(y) = \int f(z) f_{u_1|X_1=x}((z-y)/\theta_0) dz \; .$$

The transition density, denoted by $p(x, y)$, i.e., the density of $X_t|X_{t-1} = x$, will be $p(x, y) = f_x(y - g(x))$. Its transition law will be denoted by $P(x, \cdot)$ while $P^n(x, \cdot)$ stands for the law of $X_t|X_{t-n} = x$.

In order to prove ergodicity and geometric ergodicity we will use similar techniques to those of Mokkadem (1987).

2 Ergodicity of NARMA(1,1) Processes

Assume that:

H1. g is bounded over compact sets.

H2. $\inf_{x \in K_1 u \in K_2} f_x(u) > b(K_1, K_2) > 0$ for all compact sets K_1 and K_2.

H3. There exist $M > 0$ and $\eta > E\,(|u_1|)$ such that

(i) $|g(x)| + |\theta_0| r^+(x) \leq |x| - \eta$ for $|x| > M$

(ii) $\sup_{|x| \leq M} r^+(x) < \infty$.

with $r^+(x) = E(|u_1| \mid X_1 = x)$.

Note that **H1** and **H3** (ii) entail that $\sup_{|x| \leq M} \left[|g(x)| + |\theta_0| r^+(x) \right] < \infty$.

Remark 2.1. It is easy to see that similar arguments to those used in Proposition 1 of Mokkaden (1987) (using the ergodicity criterium given by Tweedie (1975)) reduce the problem of proving ergodicity to show the following conditions:

A1. For all Borelian set A with $\lambda(A) \neq 0$ and any compact set $K \subset \mathbb{R}$, there exists a positive integer n_0 such that

$$\inf_{x \in K} P^{n_0}(x, A) > 0 \ .$$

A2. There exist $M > 0$, $\eta > 0$ and $s > 0$ such that

$$E|g(x) + \varepsilon_{x,t}|^s \leq |x|^s - \eta \text{ for } |x| > M \text{ and } \sup_{|x| \leq M} E|g(x) + \varepsilon_{x,t}|^s < \infty,$$

where $\varepsilon_{x,t}$ is a random variable with distribution given by the law of $\varepsilon_t | X_{t-1} = x$;

while aperiodicity is implied by condition

A3. There exists $n_1 \in \mathbb{N}$ such that $P^{n_1}(x, \cdot)$ and λ are equivalent for all x.

Proposition 2.1. *Under* **H1** *and* **H2**, *conditions* **A1** *and* **A3** *are fulfilled.*

Proof. a) We begin by showing that **A1** holds. Let A be a Borelian set such that $\lambda(A) > 0$ and K a compact set. Since there exists a bounded set $B \subset A$ such that $\lambda(B) > 0$ we have:

$$P(x, A) = \int_A p(x, y) dy \geq \int_B f_x(y - g(x)) dy = \int_{B - g(x)} f_x(u) du \ .$$

Let $C = \bigcup_{x \in K} (B - g(x)) \subset \overline{\bigcup_{x \in K} (B - g(x))} = K^*$, K^* is a compact set since g is bounded on K. Therefore

$$P(x, A) \geq b(K, K^*) \ \lambda(B - g(x)) = b(K, K^*) \lambda(B) > 0$$

and **A1** holds with $n_0 = 1$.

b) We shall now see that **A3** holds for $n_1 = 1$. If $\lambda(A) = 0$ then $P(x, A) = \int_A p(x, y) dy = 0$ for all x. On the other hand since $f_x(u) > 0$, if $P(x, A) = 0$ for all x, we have $\lambda(A) = 0$. □

Remark 2.2. Since P^n is absolutely continuous with respect to λ, under **A1** the chain is strongly irreducible.

Let π be a sub-invariant measure for $\{X_t\}$; in the ergodic case, π is the invariant probability. As in Lemma 1 of Mokkaden (1987) we have that under **H1** and **H2** for each compact set $K \subset \mathbb{R}$ $\lambda(K) > 0$ implies $0 < \pi(K) < \infty$. Thus, in order to obtain the ergodicity of the process defined by (1.1) it remains to prove **A2**.

Proposition 2.2. *Under* **H1** , **H2** *and* **H3** *any* NARMA *(1,1) process is ergodic.*

Proof. Ergodicity follows from Remark 2.1, using Proposition 2.1 and the fact that **A2** follows easily from **H3** and the following inequality

$$E(|g(x) + \varepsilon_{x,t}|) \leq |g(x)| + E(|u_1|) + |\theta_0| r^+(x) \ . \qquad \square$$

Remark 2.3. As in Mokkaden (1987, Proposition 2) **H1** , **H2** and the following condition

C1. There exist $M > 0$ and $\eta > 0$ such that

$$|g(x)| \leq |x| - \eta \quad \text{for} \quad |x| > M \ ,$$

will entail ergodicity if $E(\varepsilon_{x,t}) = 0$. However, this last condition will not be verified by NARMA models in most of the situations. For instance, in the linear case with zero mean normally distributed errors $E(u_{t-1} \mid X_{t-1} = x) \neq 0$ and therefore $E(\varepsilon_{x,t}) \neq 0$.

3 Geometric Ergodicity and Mixing Properties for NARMA Processes

We recall the following definition and results:

- A Markov chain $\{X_t\}$ is geometrically ergodic if there exists $0 < \rho < 1$ such that $\|P^n(x,\cdot) - \pi\| = O(\rho^n)$ for almost all $x(\pi)$, where $\|\cdot\|$ stands for the total variation norm.
- In Nummelin and Tuominen (1982) it is shown that if $\{X_t\}$ is geometrically ergodic Harris recurrent and aperiodic then

$$\int \|P^n(x,\cdot) - \pi\| \pi(dx) = O(\rho^n) \ . \qquad (4)$$

- Finally in Rosenblatt (1971) it is shown that (4) implies that the process $\{X_t\}$ is α-mixing with $\alpha(n) = a^n$, for some $0 < a < 1$ (geometrically α-mixing process).

Proposition 3.1. *Under* **H1** , **H2** *and* **H3** *the chain* $\{X_t\}$ *defined by (3) is Harris-recurrent and* π *and* λ *are equivalent.*

Proof. Since **H1** and **H2** entail **A1** the process $\{X_t\}$ is strongly-irreducible (see Tweedie (1976)). On the other hand by Proposition 2.2, **H1** , **H2** and **H3** imply the ergodicity of $\{X_t\}$ and therefore the conclusion follows from a result of Tweedie (1976). $\square$

Again as in Mokkaden (1987), **A1**, **A2** and the following condition:

A4. There exist $s > 0$, $M > 0$ and $0 < \rho < 1$ such that

$$E(|g(x) + \varepsilon_{x,t}|^s) \le \rho |x|^s \quad \text{for} \quad |x| > M$$
$$\sup_{|x| \le M} E(|g(x) + \varepsilon_{x,t}|^s) < \infty \,,$$

implies the geometric ergodicity. Moreover π has a moment of order s.

A4 can be derived from **H1** and **H4** with

H4. (i) $r^+(x)$ is bounded over compact sets.
(ii) $|g(x)| + \theta_0 r^+(x) \le \rho|x|$ for $|x| > M$ for some $M > 0$ and $0 < \rho < 1$.

Putting all together we have the following result:

Proposition 3.2. *Under* **H1** , **H2** , **H3** *and* **H4** *any* NARMA (1,1) *process is a geometrically* α*-mixing process.*

4 Estimation in NARMA Processes

Let $\{X_t : 1 \le t \le T\}$ be observations of a stationary stochastic process satisfying the model defined through (3). In this section, we will introduce a family of estimation procedures through an iterative algorithm and we will prove their Fisher–consistency. Fisher–consistency is just a first step towards asymptotic properties.

Denote

$$r(x) = E(u_1 \mid X_1 = x)$$
$$h(x) = E(X_2 \mid X_1 = x) = g(x) - \theta_0 \, r(x) \,.$$

Thus, we have the identities

$$\begin{aligned} g(x) &= h(x) + \theta_0 r(x) \\ u_t &= (1-\theta_0 B)^{-1}(X_t - g(X_{t-1})) = (1-\theta_0 B)^{-1}\varepsilon_t \ , \end{aligned}$$

where B stands for the backward operator.

Thus, it follows easily that $g(x)$ minimizes

$$L_1(a) = E\left[(X_t + \theta_0 u_{t-1} - a)^2 \,|X_{t-1} = x\right]$$

and θ_0 minimizes

$$L_2(\theta) = E\left[\left((1-\theta B)^{-1}\varepsilon_t\right)^2\right].$$

This suggests to consider the simultaneous system of equations

$$\begin{cases} \min\limits_{a\in\mathbb{R}} E\left[(X_t + \theta u_{t-1} - a)^2 \,|X_{t-1} = x\right] = E\left[(X_t + \theta u_{t-1} - g_\theta(x))^2 \,|X_{t-1} = x\right] \\ \min\limits_{\theta\in(-1,1)} E\left[\left((1-\theta B)^{-1}\,(X_t - g_\theta(x))\right)^2\right] \end{cases} \tag{5}$$

It is worthwhile noting that $g_\theta(x) = h(x) + \theta r(x)$.

The following proposition shows the Fisher–consistency of the solution of the system of equations given by (5).

Proposition 4.1. *Assume that the stationary model (3) is well identified, i.e., $u_1 \neq r(X_1)$ (which holds, for instance, if (u_t, X_t) has a joint density). Then, we have that (θ_0, g) is the unique solution of (5).*

Proof. Since $g_\theta(x) = h(x) + \theta r(x)$, we have that it will be enough to show that

$$\begin{aligned} &\min_{\theta\in(-1,1)} E\left[\left((1-\theta B)^{-1}\,(X_t - h(X_{t-1}) - \theta r(X_{t-1}))\right)^2\right] \\ &\qquad = \min_{\theta\in(-1,1)} L(\theta) = L(\theta_0)\ . \end{aligned} \tag{6}$$

Note that

$$\begin{aligned} X_t - h(X_{t-1}) - \theta r(X_{t-1}) &= X_t - h(X_{t-1}) - \theta_0 r\,(X_{t-1}) + (\theta_0 - \theta)\, r\,(X_{t-1}) \\ &= X_t - g(X_{t-1}) + (\theta_0 - \theta)\, r\,(X_{t-1}) \\ &= (1-\theta_0 B)\, u_t + (\theta_0 - \theta)\, r\,(X_{t-1}) \\ &= (1-\theta B)\, u_t + (\theta_0 - \theta)\left[-Bu_t + r\,(X_{t-1})\right]\ . \end{aligned}$$

Then, if we denote by $Z_t = u_{t-1} - r\,(X_{t-1})$, we obtain

$$(1-\theta B)^{-1}\left[X_t - h\,(X_{t-1}) - \theta r\,(X_{t-1})\right] = u_t + (\theta-\theta_0)(1-\theta B)^{-1} Z_t\ ,$$

which implies that

$$L(\theta) = E\left(u_t^2\right) + (\theta-\theta_0)^2\, E\left[\left((1-\theta B)^{-1} Z_t\right)^2\right],$$

since u_t is independent of $\{X_{t-j} : j \geq 1\}$.

Now, (6) follows easily using that $P\,(u_t - r(X_t) = 0) \neq 1$ which entails that

$$E\left[\left((1-\theta B)^{-1} Z_t\right)^2\right] > 0$$

and thus $L(\theta) > L(\theta_0)$ for $\theta \neq \theta_0$.

4.1 Iterative procedure

The system of equations (5) suggest the following iterative procedure to estimate the function g and the parameter θ.

Given an initial estimate, $g^{(1)}(x)$ of $g(x)$, define the estimators through the following procedure

i) Define

$$\widehat{u}_t^{(1)} = \widehat{u}_t^{(1)}(\theta) = (1-\theta B)^{-1}\left(X_t - g^{(1)}\,(X_{t-1})\right) \tag{7}$$

and $\theta^{(1)} = \arg\min_{\theta\in(-1,1)} L^{(1)}(\theta)$ where $L^{(1)}(\theta) = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T}\left(\widehat{u}_t^{(1)}(\theta)\right)^2$.

ii) Given $\theta^{(j)}$, the estimator $g^{(j+1)}$ is defined as the nonparametric autoregression estimate

$$g^{(j+1)}(x) = \sum_{t=1}^{T-1} w_{tT}(x)\left(X_{t+1} + \theta^{(j)}\widehat{u}_t^{(j)}(\theta^{(j)})\right)\ , \tag{8}$$

where the local weights w_{tT} may be taken, for instance, as

$$w_{tT}(x) = \frac{K\left(\dfrac{x-X_t}{h_T}\right)}{\displaystyle\sum_{t=1}^{T-1} K\left(\dfrac{x-X_t}{h_T}\right)}\ .$$

The kernel $K : \mathbb{R} \to \mathbb{R}$ is a density function with 0 mean and finite variance and the bandwidth h_T satisfies $h_T \to 0$, $Th_T \to \infty$ as $T \to \infty$. Define

$$\widehat{u}_t^{(j+1)} = \widehat{u}_t^{(j+1)}(\theta) = (1-\theta B)^{-1}\left(X_t - g^{(j+1)}(X_{t-1})\right) \tag{9}$$

and

$$\theta^{(j+1)} = \arg\min_{\theta\in(-1,1)} L^{(j+1)}(\theta)$$

where

$$L^{(j+1)}(\theta) = \frac{1}{T}\sum_{t=1}^{T}\left(\widehat{u}_t^{(j+1)}(\theta)\right)^2.$$

Iterate until convergence. Denote $\widehat{g}$ and $\widehat{\theta}$ the resulting estimates.

Remark 4.1. Forecasting is one the important applications in ARMA models, which is typically done, after the parameters have been estimated, by using "estimated residuals" when moving averages are present. For our NARMA (1,1) model prediction may be done as follows

- For $2 \le \tau \le t-1$, define $\widehat{\epsilon}_\tau = X_\tau - \widehat{g}(X_{\tau-1})$, where $\widehat{g}$ is an estimate of the autoregression function and $\widehat{\epsilon}_\tau = 0$ otherwise.
- Given $\widehat{\theta}$ an estimate of the moving average parameter, let $\widehat{u}_\tau = \left(1-\widehat{\theta}B\right)^{-1}\widehat{\epsilon}_\tau$
- Predict the observation at time t as $\widehat{g}(X_{t-1}) - \widehat{\theta}\,\widehat{u}_{t-1}$

4.2 An example

We will illustrate the iterative procedure through two simulated examples.

We have simulated two series following the model

$$X_t = g(X_{t-1}) + u_t - \theta_0 u_{t-1}$$

where the random variables $u_t \sim N(0,\sigma^2)$ are independent and identically distributed; and u_t is independent of $\{X_{t-1}, X_{t-2}, \dots\}$.

In the first case, referred as a) in Figure 1, $g(x) = 0.5x$ and $\theta = -0.1$ while in the second one $g(x) = 0.5\min(|x|, 1.345)\mathrm{sg}(x)$ and $\theta = -0.6$. In all cases, we have used a Gaussian kernel with $\sigma = 0.37037$ and a bandwidth $h_T = 0.5$. For each data set we have three plots: a_1) is a plot of X_t against X_{t-1}; a_2) is a plot of the series X_t while in a_3) we plot the function $g(x)$. The sample size is 1000. The next three plots are the corresponding plots for the second model (b_1) to b_3)). While the plots in a_1) and a_2) are very similar to those in b_1) and b_2) the true underlying g functions are different.

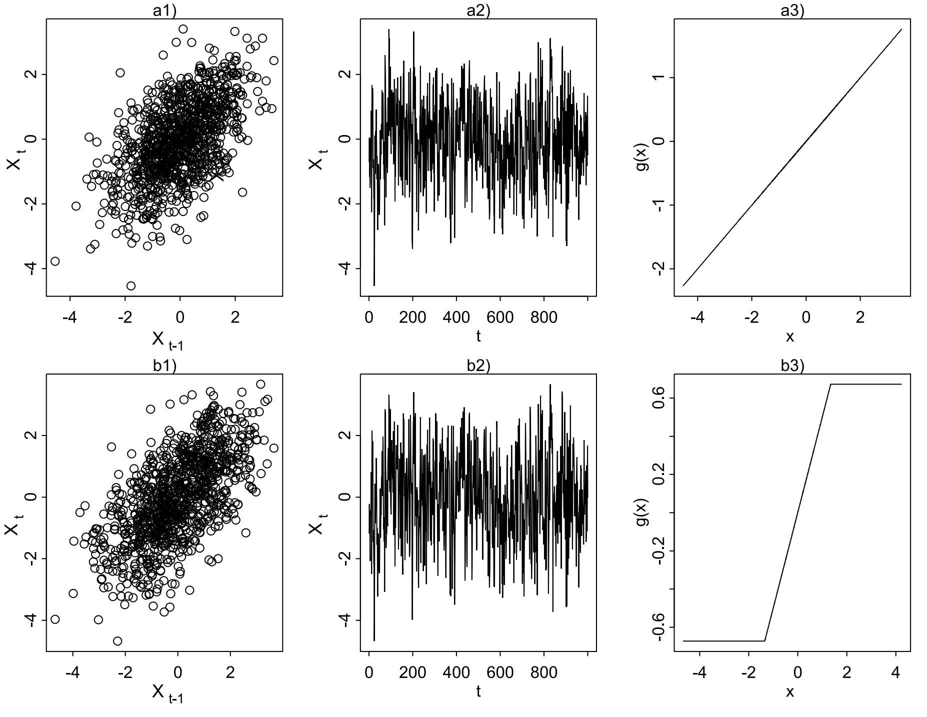

Figure 1: Generated data sets.

Figure 2 gives the estimated autoregression function for the first two steps and the last one. For both models, in the first step, an ARMA (1,1) model was fitted to provide the initial estimates of the autoregression function through the autoregressive fitted parameter. For the linear model a) $\hat{\phi} = 0.4999$, while $\hat{\phi} = 0.3043$ for model b). Thus, an ARMA fits the autoregression function as $g^{(1)} = 0.4999x$ for the linear model (true ARMA model), while it fits the autoregression function as $g^{(1)} = 0.3043x$ in the nonlinear case. As expected in the linear case the fit is accurate while in the nonlinear case, even the slope at the central part (0.5) is estimated by 0.3043. The plots given in Figure 2 show the improvement obtained in the estimation by the nonparametric procedure we have introduced. Even in the linear case, the smoother does quite well. Table 1 gives the values of $\theta^{(j)}$ for different values of j.

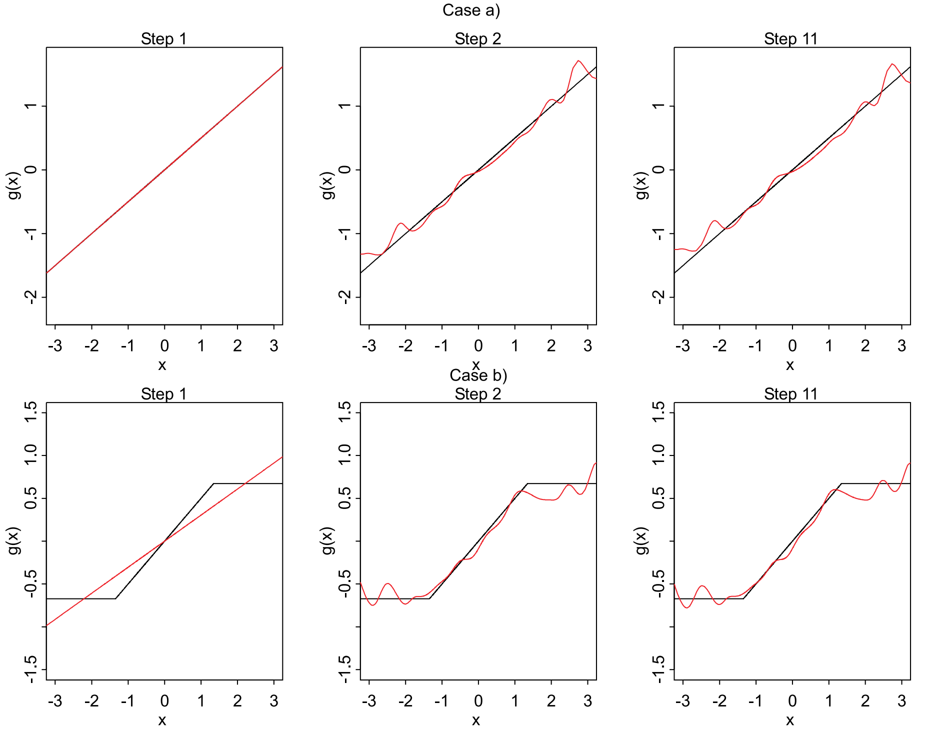

Figure 2: Estimates of the autoregression function.

Step	1	2	3	4
Case a)	-0.08315	-0.09352	-0.10032	-0.10488
Case b)	-0.58908	-0.59867	-0.59739	-0.59619

Step	8	9	10	11
Case a)	-0.11237	-0.11301	-0.11343	-0.11372
Case b)	-0.59528	-0.59526	-0.59525	-0.59525

Table 1: Estimates for the parameter θ.

Acknowledgment. The authors wish thank an anonymous referee for valuable comments which led to an improved version of the original paper.

References

Bosq, D. *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Lectures Notes in Statistics, **110** (1996), Springer.

Carbon, M. and Delecroix, M. Nonparametric forecasting in time series, a computational point of view. *Applied Stoch. Models and Data Analysis* **9** (1993), 215–229.

Granger, C. and Newbold, P. *Forecasting Economic Times Series*. New York: Academic Press, (1986).

Mokkadem, A. Sur un modèle autorégressif non linéaire, ergodicité et ergodicité géométrique. *J. Time Ser. Anal.* **2** (1987), 195–204.

Nummelin, E. and Tuominen, P. Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory. *Stoch. Proc. Appl.* **2** (1982), 187–202.

Rosenblatt, M. *Markov processes. Structure and asymptotic behaviour*. Berlin, Springer-Verlag, (1971).

Tweedie, R.L. Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. *Stoch. Proc. Appl.* **3** (1975), 385–403.

Tweedie, R.L. Criteria for classifying general Markov chains. *Adv. Appl. Prob.* **8** (1976), 737–771.

Graciela Boente
Universidad de Buenos Aires and CONICET,
Buenos Aires
ARGENTINA

Ricardo Fraiman
Universidad de la República,
Montevideo
URUGUAY