Statistical Methods & Applications (2004) 13: 167–178
DOI: 10.1007/s10260-004-0096-0

SMA

The univariate distribution function for a particular bilinear model

Francesco Giordano

Dipartimento di Scienze Economiche e Statistiche, Università di Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy (e-mail: giordano@unisa.it)

Received February 2003 / Revised May 2004

Abstract. The aim of the paper is to find the univariate stationary distribution of a particular bilinear process. In this context, we propose a novel approach to derive the distribution function. It is based on a recursive formula and allows to relax the conditions on the moments of the process. We also show that the derived approximation converges to the "true" distribution function. The accuracy of the recursive formula is evaluated for finite sample dimensions by a small simulation study.

Key words: Nonlinearity, Bilinear Model, Edgeworth series

1 Introduction

In recent years the bilinear models for time series have gained a growing attention. These models are difficult to deal with because of their complex probabilistic structure (Tong, 1990). Therefore we focus on a particular bilinear model which is often used as a building block for much more complex non linear time series models such as the GARCH-BL (Storti, 2003).

Consider the following bilinear model:

$$X_t = \epsilon_t + b\epsilon_{t-1}X_{t-2} \tag{1}$$

where b is an unknown parameter, $\epsilon_t \sim N(0, \sigma^2)$, *iid*, and, without loss of generality, $\sigma^2 = 1$.

This specification is the simplest case of Type I class of Standardized bilinear models (Grahn, 1995). It can be easily confused with a White Noise because it has zero mean, constant variance and zero autocovariances. But it has a more complex structure in terms of the third mixed moments. Furthermore the time series generated by this process show peaks in the observations which can be very distant from the stationary level of the series. By analyzing the series with linear models, these peaks look like outlying observations. However the stationary univariate distribution of X_t in (1) is different from the Normal one (assigned to the residuals ϵ_t).

The aim of this paper is to derive an approximation of the stationary distribution of X_t in (1). In particular, we propose two different approaches: the first is based on a type of Edgeworth expansion while the second is obtained by using a recursive formula. We show that the last approximation is better than both the Normal and Edgeworth type approximations.

The paper is organized as follows. In Sect. 2 we present and discuss some preliminary results and derive the Edgeworth type approximation. In Sect. 3 we propose an alternative approach based on a recursive formula and we show that, under very weak conditions, it gives the best approximation. In order to evaluate the previous results, a small simulation experiment is reported in Sect. 4. Some concluding remarks are given in the final section.

2 Preliminary results and the Edgeworth type approximation

The following lemma is needed to introduce the Edgeworth type expansion for the distribution function of X_t.

Lemma 1 *If X_t is defined as in (1) and X_t is a strong stationary process, then its distribution function must be absolutely continuous with density function $f(.;b)$ symmetric, positive, continuous and unimodal.*

Proof. The process X_t is stationary if $|b| < 1$ (Grahn, 1995).

Given model (1) and being X_t a strong stationary process, it follows that its characteristic function, $H(u;b)$, is:

$$H(u;b) = e^{-\frac{1}{2}u^2} \int_{\mathcal{R}} e^{-\frac{1}{2}b^2u^2y^2} dF(y;b) \tag{2}$$

where $F(.;b)$ is the stationary distribution function of X_t.

Furthermore $H(u;b) = 1$ if and only if $u = 0$. Then the distribution function cannot be arithmetic at some points over $\mathcal{R}$ and so it is a continuous distribution function.

Being $e^{-\frac{1}{2}b^2u^2y^2}$ uniformly bounded, and since

$$\int_{\mathcal{R}} |H(u;b)|\, du = \sqrt{2\pi} \int_{\mathcal{R}} \frac{1}{\sqrt{1+b^2y^2}} dF(y;b) \leq \sqrt{2\pi} < \infty$$

then, there exists a continuous density function $f(.;b)$.

Moreover the characteristic function is real and so $f(.;b)$ is symmetric.

By using the Inversion Theorem, $f(.;b)$ is

$$f(x;b) = \int_{\mathcal{R}} \frac{1}{\sqrt{2\pi\,(1+b^2y^2)}} e^{-\frac{x^2}{2(1+b^2y^2)}} f(y;b)dy \tag{3}$$

and so $f(.;b)$ is defined by a second kind integral equation.

The density function is always positive. In fact, if $f(x_0; b) = 0$ then

$$\int_{\mathcal{R}} \frac{1}{\sqrt{2\pi\left(1+b^2y^2\right)}} e^{-\frac{x_0^2}{2(1+b^2y^2)}} f(y; b)dy = 0.$$

The last equation can be satisfied only for $f(x; b) = 0, \forall x \in \mathcal{R}$, which is impossible.

Given that the integrand function in (3) is:

$$g(x, y; b) = \frac{1}{\sqrt{2\pi\left(1+b^2y^2\right)}} e^{-\frac{x^2}{2(1+b^2y^2)}}$$

has the first partial derivative with respect to x, and it is uniformly dominated by $1/\sqrt{2\pi}$, then

$$\frac{df(x; b)}{dx} = -x \int_{\mathcal{R}} \frac{1}{\sqrt{2\pi\left(1+b^2y^2\right)^3}} e^{-\frac{x^2}{2(1+b^2y^2)}} f(y; b)dy$$

So $f(.; b)$ has a maximum only at $x = 0$.

It follows that $f(.; b)$ is unimodal. □

By Lemma 1 the expression (3) is the true univariate density function for X_t in (1). However this lemma does not give a way to obtain a solution for (3). A series expansion can be used to approximate the true distribution function for X_t in (1). So it can be stated the following theorem.

Theorem 1 *If X_t is defined as in (1), X_t is a strong stationary process and $|b| < \sqrt{2/5}$, then there exists a series approximation of the univariate distribution function at the second term*

$$\tilde{F}(x) = \Phi(x) - cx\phi(x) \tag{4}$$

with $c = \frac{b^2}{2(1-b^2)}$, $\phi(x)$ and $\Phi(x)$ the standard Normal density and distribution functions, respectively.

Besides, the approximation error is

$$\int_{-\infty}^{\infty} \left[f(x) - \tilde{f}(x)\right]^2 dx \leq \frac{105}{64} b^8 (\mu_4)^2 \sqrt{\pi}$$

with $f(\cdot)$ the true density function, $\tilde{f}(\cdot)$ the density function corresponding to $\tilde{F}(\cdot)$ and $\mu_4 = E(X_t^4)$.

Proof. Starting from (2) and, by Taylor expansion at the second term, we have an approximate characteristic function $\tilde{H}(u; b)$,

$$\tilde{H}(u; b) = e^{-\frac{1}{2}u^2} \int_{\mathcal{R}} \left(1 - \frac{1}{2} b^2 u^2 y^2\right) f(y; b)dy \tag{5}$$

By the Inversion Theorem we have the result in (4). The expression (4) has the density function of Lemma 1 if $|b| < \sqrt{2/5}$. If this condition is violated the density function, corresponding to $\tilde{F}(x)$ in (4), is not unimodal.

Let $R_1(u;b)$ be the remainder term of the expansion (5) which is

$$R_1(u;b) \leq \frac{1}{8}u^4 b^4 \mu_4 e^{-\frac{u^2}{2}}$$

By Parseval identity and, again, by the Inversion Theorem, we have

$$\int_{-\infty}^{\infty}\left[f(x)-\tilde{f}(x)\right]^2 dx \leq \frac{105}{64} b^8 (\mu_4)^2 \sqrt{\pi}$$

since $\int_{-\infty}^{\infty} u^8 e^{-u^2} = \sqrt{\pi}105$. □

Remark 1 $|b| < \sqrt{2/5}$ implies that the sixth moment exists, that is $E(X_t^6) < \infty$.

Remark 2 The distribution (4) can be seen as an Edgeworth series truncated at the second term. So Theorem 1 states that there is an Edgeworth expansion for the stationary univariate distribution function for X_t defined in (1).

The above theorem gives a general method to approximate the unknown distribution function. It is based on the existence of the moments of X_t. Since in model (1) ϵ_t is a Gaussian White Noise then (Grahn, 1995)

$$E(X_t^n) < \infty \Leftrightarrow b^n < \frac{1}{1\cdot 3 \cdots (n-1)}$$

for n even.

So it implies that a better approximation imposes the existence of higher order moments which means that the parameter b tends to zero or equivalently the non linear structure in model (1) disappears.

Moreover it is possible to find a class of approximate distribution functions of X_t without imposing conditions on the existence of the moments. So we can compare the approximate distribution function of Theorem 1 with the distribution functions shown in the following pages.

3 An alternative approach to approximate the distribution function

Firstly, we introduce a proper space of distribution functions.

Definition Let $\mathcal{G}$ be the class of functions $g(.;b)$ over $\mathcal{R}$ such that $g(.;b)$ is continuous, unimodal, symmetric density function with $\int_{\mathcal{R}} xg(x;b)dx = 0$ and $\mu_2 = E(X_t^2) = \int_{\mathcal{R}} x^2 g(x;b)dx = \frac{1}{1-b^2}$.

In the following we use L^1 to denote the class of functions which are absolutely integrable over $\mathcal{R}$ and L^2 to indicate the class of functions which are square integrable over $\mathcal{R}$ and $\|.\|_{L^k}$, $k=1,2$, as the relative norm on L^k, $k=1,2$.

Lemma 2 *The class $\mathcal{G} \subset L^2$ and it is complete in the L^2 norm.*

Proof. Given a Cauchy sequence in $\mathcal{G}$, $g_1, \dots, g_n, \dots$ then $\forall \epsilon > 0 \quad \exists n_0 :$

$$\forall (m, n) \quad m > n > n_0 \quad \left| \|g_m\|_{L^2}^2 - \|g_n\|_{L^2}^2 \right| \le \|g_m - g_n\|_{L^2}^2 < \epsilon$$

So the sequence $\|g_n\|_{L^2}^2$ is convergent and there exists a function $\bar{g}$ such that

$$\|g_n\|_{L^2}^2 \to \|\bar{g}\|_{L^2}^2 < \infty$$

So $\mathcal{G} \subset L^2$.

But $\|g_n\|_{L^1} = 1 \quad \forall n$ and $\|g_n\|_{L^1} \to \|\bar{g}\|_{L^1} = 1$. Therefore $\bar{g}$ is a density function.

Every function $g_n(x;b)$ is bounded and continuous with respect to x, $\forall n$ and $\forall |b| < 1$. Consider a compact subset $K = [-q, q] \subset \mathcal{R}$ with q a positive and finite number. Then we can build the functions g_n^K in the following way:

$$g_n^K(x;b) = \begin{cases} g_n(x;b) & x \in K \\ 0 & x \notin K \end{cases}$$

The functions g_n^K are dense in $\mathcal{G}$. Then fix $\epsilon > 0 \quad m', n' \quad m' > n' > n_0$ such that

$$\int_K \left(g_{m'}^K(x;b) - g_{n'}^K(x;b) \right)^2 dx < \epsilon$$

By the mean value theorem there exists a $x_0 \in K$ such that

$$\left| g_{m'}^K(x_0;b) - g_{n'}^K(x_0;b) \right| < \sqrt{2q\epsilon} = \epsilon'$$

Also, the choice of x_0 depends on m and n. Since g_n^K is bounded then $\forall (m, n)$ we can find a value of x such that

$$\left| g_m^K(x;b) - g_n^K(x;b) \right| < \epsilon'$$

Hence the sequence g_n^K converges uniformly to $\bar{g}^K$. Finally also the sequence g_n converges uniformly to $\bar{g}$. Then $\bar{g}$ is bounded and continuous.

Since $\int_{\mathcal{R}} x^{2k-1} g_n(x;b) dx = 0 \quad \forall n, \quad$ then $\int_{\mathcal{R}} x^{2k-1} \bar{g}(x;b) dx = 0 \quad k \in \mathcal{N}$.

It follows that $\bar{g}$ is symmetric.

In the same way $\int_{\mathcal{R}} x^2 g_n(x;b) dx \to \int_{\mathcal{R}} x^2 \bar{g}(x;b) dx = \mu_2^X$.

$\bar{g}$ is unimodal. In fact, suppose $\bar{g}$ is not unimodal. If $\bar{g}$ has a maximum at $x_0 > 0$ then there exist $x' < x''$ such that $\bar{g}(x';b) = \bar{g}(x'';b)$. Then

$$\lim g_n(x';b) = \bar{g}(x';b) \quad \lim g_n(x'';b) = \bar{g}(x'';b)$$

So $\lim |g_n(x';b) - g_n(x'';b)| = 0$. Fix $\epsilon > 0 \quad \exists n_0 : \forall n > n_0$ we have

$$|g_n(x';b) - g_n(x'';b)| < \epsilon \Rightarrow g_n(x';b) = g_n(x'';b)$$

But this is impossible because g_n is a bounded, continuous and unimodal function with the only maximum at 0. $\square$

Remark The true density function, f, of the process (1) and the approximation $\tilde{f}$ belong to $\mathcal{G}$, where $\tilde{f}$ is derived by (4) in Theorem 1. In fact, it is $\tilde{f}(x;b) = (1-c)\phi(x) + cx^2\phi(x)$.

To understand the degree of approximation of $\tilde{f}$ we can choose another distribution function which belongs to $\mathcal{G}$. Such a distribution is the Normal one with zero mean and variance μ_2. Let

$$p(x;b) = \frac{1}{\sqrt{2\pi\mu_2}} e^{-\frac{x^2}{2\mu_2}}$$

Lemma 3 *If* $|b| < \sqrt{2/5}$ *then* p *is a better approximation than* $\tilde{f}$ *for* f *in* L^2 *norm, that is*

$$\|p - f\|_{L^2} < \|\tilde{f} - f\|_{L^2}.$$

Proof. By Parseval identity we can evaluate the characteristic functions. Let $H(u;b)$ be the characteristic function of f given by expression (2) in Lemma 1, $\tilde{H}(u;b)$ be the characteristic function of $\tilde{f}$ given by expression (5) in Theorem 1 and $H_p(u;b)$ be the characteristic function of p. The condition $|b| < \sqrt{2/5}$ assures that $\tilde{f}$ exists. So

$$H(u;b) = \tilde{H}(u;b) + R_1(u;b)$$

where $R_1(u;b) < \frac{1}{8}u^4 b^4 \mu_4 e^{-\frac{u^2}{2}}$ with $\mu_4 = E(X_t^4)$ as in Theorem 1.

But also

$$H_p(u;b) = \exp\left(-\frac{u^2}{2}\right)\exp\left(-\frac{u^2 b^2 \mu_2}{2}\right) = \tilde{H}(u;b) + R_2(u;b)$$

where $R_2(u;b) = \frac{1}{8}u^4 b^4 (\mu_2)^2 e^{-\frac{u^2}{2}} \alpha_{u,b}$, with $0 < \alpha_{u,b} \leq 1$ and $\alpha_{u,b}$ depends on u and b. So

$$\|\tilde{H} - H\|_{L^2} \leq \|R_1\|_{L^2} \quad \text{and} \quad \|H_p - H\|_{L^2} \leq \|R_1 - R_2\|_{L^2}$$

Since $R_1(u;b) > R_2(u;b) \forall u, b$ given that $\mu_4 > (\mu_2)^2 \alpha_{u,b}$ for the process defined in (1), it follows that

$$\|p - f\|_{L^2} < \|\tilde{f} - f\|_{L^2} \qquad \square$$

Remark Observe that the Normal distribution, with density function p, approximates (in L^2 norm) the true density function f better than $\tilde{f}$. Moreover p is defined for $|b| < 1$ whereas $\tilde{f}$ for $|b| < \sqrt{2/5}$.

To build an approximation for f, $\forall |b| < 1$, which does not suffer the limitation of $\tilde{f}$, we use the second kind integral equation of Lemma 1 as recursive formula, that is

$$f_{k+1}(x;b) = \int_{\Re} \frac{1}{\sqrt{2\pi\left(1 + b^2 y^2\right)}} e^{-\frac{x^2}{2(1+b^2y^2)}} f_k(y;b)\,dy$$

with $k = 0, 1, 2, \ldots$ and $f_0 \in \mathcal{G}$.

The above equation can be written as

$$f_{k+1} = V(f_k) \tag{6}$$

and by Lemma 1 it is easy to show that V is an operator from $\mathcal{G}$ to $\mathcal{G}$.

Lemma 4 *The sequence $\{f_k\}$, as defined in (6), converges to f in L^2 norm, $\forall f_0 \in \mathcal{G}$.*

Proof. Since $\|f_k\|_{L^1} = 1, \quad \forall k$ then $\|V\|_{L^1} = 1$. By Lemma 1 it can be argued that $f_k(x;b) < 1 \quad \forall(k, x, b)$. So the kernel of the operator V is always positive and less than 1. It follows that $\|V\|_{L^2} \le \|V\|_{L^1} = 1$. Then the operator V is continuous and the sequence $\|f_k\|_{L^2}$ is bounded.

The sequence $\|f_k\|_{L^2}$, or equivalently $\|f_k\|^2_{L^2}$, is monotone. Suppose that the monotonicity does not hold. There exists, at least a k, for which $\|f_k\|^2_{L^2} \ge \|f_{k+1}\|^2_{L^2} < \|f_{k+2}\|^2_{L^2}$.

The above relations can be written as

$$\int (f_k(x;b) - f_{k+1}(x;b))\,(f_k(x;b) + f_{k+1}(x;b))dx \ge 0$$

$$\int (f_{k+2}(x;b) - f_{k+1}(x;b))\,(f_{k+2}(x;b) + f_{k+1}(x;b))dx > 0$$

Since $(f_k(x;b) + f_{k+1}(x;b))$ and $(f_{k+2}(x;b) + f_{k+1}(x;b))$ are also quantities less than 1 by Lemma 1, the second expression becomes

$$\int (f_{k+2}(x;b) - f_{k+1}(x;b))\,(f_{k+2}(x;b) + f_{k+1}(x;b))dx \le$$

$$\le \int (f_{k+2}(x;b) - f_{k+1}(x;b))dx = 0$$

But this is impossible and so we can state that the sequence $\|f_k\|^2_{L^2}$, or equivalently $\|f_k\|_{L^2}$, is monotone.

Thus, this sequence is convergent and so there exists a function $\bar{g} \in L^2$, with $\|\bar{g}\|_{L^1} = 1$, such that $\|f_k\|_{L^2} \to \|\bar{g}\|_{L^2}$.

By Lemma 1, the characteristic functions H_k are integrable, $\forall k$, so $\bar{g}$ is a bounded and continuous function over $\mathcal{R}$. It follows that the sequence $\{f_k(x;b)\}$ is uniformly convergent with respect to x. Therefore

$$\lim_k \int_{\mathcal{R}} |f_k(x;b) - \bar{g}(x;b)|^2\,dx \quad = \quad \int \lim_k |f_k(x;b) - \bar{g}(x;b)|^2\,dx = 0.$$

Hence f_k converges to $\bar{g}$ in L^2 norm, that is $\|f_k - \bar{g}\|_{L^2} \to 0$. If f_0, the initial point of the sequence in (6), is a generic function which belongs to $\mathcal{G}$, since $f_k \in \mathcal{G}$ and $\mathcal{G}$ is complete in L^2 norm by Lemma 2, it follows that $\bar{g} \in \mathcal{G}$.

Finally, if f is a solution of

$$f = V(f)$$

then $f = \bar{g}$ in L^2 norm. □

An explicit expression can be derived for the characteristic function, H_1, corresponding to the distribution function with density $f_1 = V(p)$. By (2) in Lemma 1 H_1 is

$$H_1(u; b) = e^{-\frac{u^2}{2}} \left(1 + b^2 \mu_2 u^2\right)^{-1/2}. \tag{7}$$

Corollary 1 *If f_0 in the sequence (6) is p, the Normal distribution, then $f_1 = V(p)$ is a better approximation than p for f, in L^2 norm.*

Proof. Since $p \in \mathcal{G}$ then $\|p - f\|_{L^2} \geq \|f_1 - f\|_{L^2}$ by Lemma 4.

It is easy to verify that

$$\|H_p\|_{L^2} < \|H_1\|_{L^2}$$

where H_p is the characteristic function of p.

Therefore, by Parseval identity

$$\|p - f\|_{L^2} > \|f_1 - f\|_{L^2}.$$ □

Lemma 4 assures that formula (6) is convergent to f for every initial point in the class $\mathcal{G}$. So we have a method to build an approximation to f without imposing any further condition on the moments of X_t. Only the second moment must exist because the class $\mathcal{G}$ has distribution functions with a finite second moment, by definition. In fact, it easy to verify that the existence of the second moment assures the stationarity of X_t defined in (1) (Grahn, 1995). But the main difficulty to go on in the recursion formula (6) is to derive analytically the density functions $f_k(\cdot)$.

At this point we have three distribution functions, $\tilde{f}$ by Theorem 1, p the Normal distribution and f_1 by corollary 1 which approximate f. The meaning of these approximations can be analyzed with respect to a class of stationary stochastic processes approximating the bilinear one defined in (1). For this purpose we can see such a class of stationary stochastic processes as

$$Z_t = \epsilon_t + m(\epsilon_{t-1}) \tag{8}$$

or

$$Z'_t = T(Z'_{t-1}) + \epsilon_t \tag{9}$$

where ϵ_t is a Gaussian White Noise as in (1) and m and T are strictly monotone and continuously differentiable functions over $\mathcal{R}$.

The following theorem shows a link between the density functions $f_k(\cdot)$ and the class of processes (8) and (9).

Theorem 2 *In the class of processes in (8) and (9), with density defined in (6), there exists only one linear process as approximation of the bilinear one given in (1).*

Proof. Let H_Z and $H_{Z'}$ be the characteristic functions for Z_t and Z'_t, respectively. By corollary 1, we can start with $H_Z \equiv H_p$ corresponding to the Normal distribution function in $\mathcal{G}$. In this case it is easy to verify that the process Z_t becomes a $MA(1)$ with coefficient $b\sqrt{\mu_2}$. In a similar way we have an $AR(1)$ process for Z'_t.

Now consider a $k \geq 1$. So $H_Z \equiv H_k$, where H_k is the characteristic function corresponding to a distribution function with density f_k.

Assume that k is fixed. Let $w = m(\epsilon_{t-1})$ be a continuous random variable over $\mathcal{R}$. Then, by (8)

$$H_Z(u;b) = e^{-\frac{u^2}{2}} H_w(u;b)$$

where H_w is the characteristic function of w. By (2) in Lemma 1, the characteristic function of H_Z can always be written as

$$H_Z(u;b) = e^{-\frac{u^2}{2}} (h_Z)_k(u;b)$$

where $H_w(u;b) = (h_Z)_k(u;b) = \int_{\mathcal{R}} e^{-\frac{1}{2}b^2u^2y^2} f_{k-1}(y;b)dy$.

But $\|(h_Z)_k\|_{L^1} = \infty, \forall k \geq 1$. Further H_w is always non negative, then the random variable w has no bounded density function over $\mathcal{R}$. It follows that $m'(x) = 0$ for some finite x_0. Thus m is not strictly monotone as required by definition of m in (8).

The same result is true for Z'_t in (9). $\square$

Remark The above theorem states that only a linear process can be used to approximate the bilinear one in (1) of class (8) or (9). Nevertheless a non linear process, in the same classes, exists as competitor of (1) if the functions m or T are restricted to a subset of $\mathcal{R}$.

The method of series expansion in Section 2 implies that it is necessary to restrict m or T to a compact subset of $\mathcal{R}$ because the Fourier Transform of the function $exp\left(\frac{1}{2}u^2\right)\tilde{H}(u;b)$ exists only in a compact subset of $\mathcal{R}$. $\tilde{f}$ is an example of this case. Instead, the other approach with the recursive formula (6), allows to use a non linear process (8) or (9) with the functions m or T restricted to $\mathcal{R} - \{x_1, x_2, \ldots, x_n\}$, where $\{x_1, x_2, \ldots, x_n\}$ is a finite set of points in $\mathcal{R}$. By corollary 1 the approximation of the density function f_1 is equivalent to a non linear function m or T which is restricted to a subset of $\mathcal{R}$ such as $\mathcal{R} - \{x_1, x_2, \ldots, x_n\}$.

The approximate density functions $\tilde{f}$, p and f_1 depend on the unknown parameter b. If we consider a consistent estimator for b, such as the (Conditional Least Squares) CLS, shown in Grahn (1995), say $\hat{b}_{CLS}$, then

$$\hat{\tilde{f}}(x) = \phi(x)(1-\hat{c}_{CLS}) + \hat{c}_{CLS}x^2\phi(x) \xrightarrow{a.s.} \tilde{f}(x).$$

This convergence can be proved by $\hat{b}_{CLS} \xrightarrow{a.s.} b$, (Grahn, 1995), and $\hat{c}_{CLS} = \frac{\hat{b}^2_{CLS}}{2(1-\hat{b}^2_{CLS})} \xrightarrow{a.s.} c$ because it is a continuous function of b.

It is also true for p and f_1.

The knowledge of a parametric distribution function allows to solve the estimation problem in a finite dimensional space instead of infinite dimensional one.

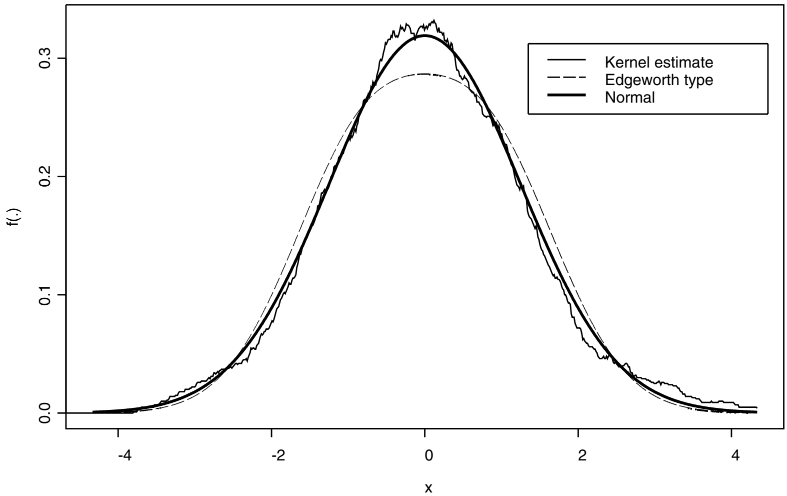

Fig. 1. Kernel estimate, Edgeworth type $\tilde{f}$ and Normal p, with $b = 0.6$

4 Empirical results

To evaluate the results of the proposed techniques a Monte Carlo study has been performed. We draw 1000 series from (1) of length T=200 observations fixing $\sigma^2 = 1$. Then, to eliminate the starting values effects, we select randomly a number between $T/2$ and $3T/4$. This number is used to extract a specific observation for each series. So we get 1000 values and show them by means a Kernel density estimation. The Kernel function is Gaussian. So the bias of the Kernel density estimation is $\frac{1}{2}h^2 f''(x) + o(h^2)$, where $f''(\cdot)$ is the second derivative of the density function $f(\cdot)$ (Jones, 1995). As in Lemma 1 it can be shown that $f''(\cdot)$ exists. Denote by h the bandwidth parameter which tends to zero when n, the number of the observations, goes to infinity. Since $n = 1000$, then the bias is a negligible quantity, $\forall x$. Also, we estimate $\hat{b}_{CLS}$ for each series and let $\bar{b}_{CLS}$ be its mean over 1000 Monte-Carlo runs.

Figure 1 reports the Kernel density estimate for the simulated values, the density function computed from the Edgeworth type $\tilde{f}$, and the Normal density function p, fixing the value of b to 0.6 which is near to $\sqrt{2/5}$ (the upper limit for $\tilde{f}$ to be a valid approximation).

It is clear that the Normal density p has a better performance with respect to $\tilde{f}$, the Edgeworth type approximation. It confirms the results in Lemma 3. For values of b closer to zero, the differences tend to disappear, and the approximation (4) works well.

Figure 2 compares the simulated density function (Kernel estimate) with the Normal density p and f_1, defined by its characteristic function in (7). The parameter b is fixed to 0.8, a value for which $\tilde{f}$ is not valid.

It is evident that f_1 is a better approximation than p. In this case it is possible to verify that f_1 has heavier tails than the Normal density p.

Since the fourth moment of (1) does not exist for $b = 0.8$, the CLS estimator of b does not satisfy the conditions of the strong consistency (Grahn, 1995).

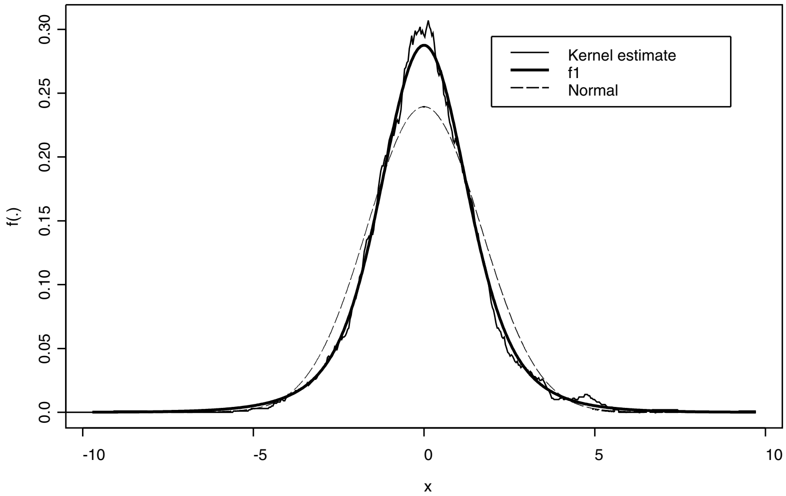

Fig. 2. Kernel estimate, f_1 and Normal p, with $b = 0.8$

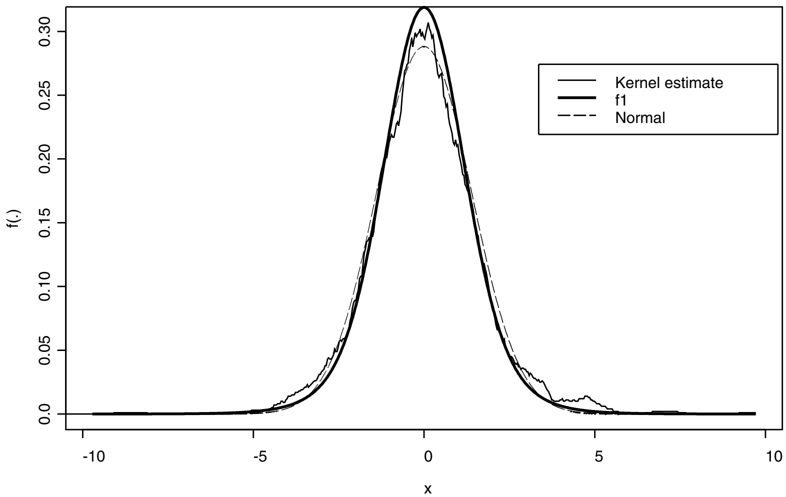

Fig. 3. Kernel estimate with the true value $b = 0.8$, f_1 and normal p with the mean of the CLS estimate $\overline{b}_{CLS} = 0.6917$

If we consider the CLS estimator for the parameter b, with $\sigma^2 = 1$, we have $\overline{b}_{CLS} = 0.5577$ for $b = 0.6$ and $\overline{b}_{CLS} = 0.6917$ for $b = 0.8$ over the 1000 Monte-Carlo runs. Since the bias is great in the latter case, in Fig. 3 we compare the true density function for $b = 0.8$ (Kernel estimate) with the functions p (Normal) and the f_1 in (7) fixing their parameter b to $\overline{b}_{CLS} = 0.6917$. So Fig. 3 shows the same true density function (Kernel estimate) as in Fig. 2. It is interesting to note as the behaviour of the density functions p and f_1 seems to approximate very well the true density function. In particular, given that $\hat{b}_{CLS}$ underestimates b for T finite (Grahn, 1995), then the approximation density function f_1 is just a good choice when we have to estimate the parameter b and it is close to 1, the boundary of non stationarity for the process defined in (1).

5 Concluding remarks

The aim of this paper is to investigate the theoretical aspects of the univariate distribution function for a particular bilinear process in (1). Two methods are proposed to derive an approximating density function. The result of the second section gives the conditions for which the Edgeworth type expansion of the univariate stationary distribution of X_t is valid. The alternative method, proposed in Sect. 3, is more general and it can be applied only assuming that X_t is stationary. Moreover the last approach delivers a parametric density function which can be useful for the estimation of the unknown parameter b (by maximum likelihood, for instance) without the constraint on the existence of higher order moments which are necessary for the applications of the techniques available at the moment.

An analytical approximating distribution of X_t allows to compute the unconditional expected length of one-step ahead prediction interval (Kabaila, 2001). In fact, in a forecasting context it is crucial to know the unconditional distribution function in order to forecast m steps ahead, especially, when $m > 1$.

The generalization of methods, which have been described in the second and third sections, to a wider class of bilinear models is still under study.

References

Brillinger D (1981) Time Series: Data Analysis and Theory. Holt, Rinehart & Winston, New York

Brockwell P, Davis R (1991) Time Series. Theory and Methods. Springer, New York

Grahn T (1995) A conditional least squares approach to bilinear time series estmation. Journal of Time Series Analysis 16, 509–529

Guegan D (1987) Different representations for bilinear models. Journal of Time Series Analysis 8, 389–408

Jones MC, Wand M (1995) Kernel Smoothing. Chapman & Hall, London

Kabaila P, He Z (2001) On prediction intervals for conditionally heteroscedastic processes. Journal of Time Series Analysis 22, 725–731

Liu S (1985) Theory of bilinear time series models. Comm. Statist. Theor. Meth. 14, 2549–2561

Storti G, Vitale C (2003) BL – GARCH models and asymmetries in volatility. Statistical Methods & Applications 12, 19–40

Subba Rao T (1981) On the theory of bilinear time series models. J. R. Statist. Soc. B 43, 244–255

Tong H (1990) Non Linear Time Series. A Dynamical System Approach. Oxford University Press, Oxford