NONPARAMETRIC REGRESSION WITH SERIALLY CORRELATED ERRORS

Jan G. De Gooijer[1], Ali Gannoun[2], and Irène Larramendy[2]

[1] Department of Economic Statistics, University of Amsterdam,
Roetersstraat 11, 1018 WB Amsterdam, The Netherlands
Telephone: +31-20-525 4244; Fax: +31-20-525 4349
e-mail: jandeg@fee.uva.nl

[2]Laboratoire de Probabilités et Statistique
Université Montpellier II
Place Eugène Bataillon, 34095 Montpellier Cédex 5, France
Telephone: +33-4-67-14-3578; +33-4-67-14-4607; Fax: +31-4-67-14-3558
e-mail: gannoun@math.univ-montp2.fr; larramen@stat.math.univ-montp2.fr

Abstract

Motivated by the problem of setting prediction intervals in time series analysis, this investigation is concerned with recovering a regression function $m(X_t)$ on the basis of noisy observations taking at random design points X_t. It is presumed that the corresponding observations are corrupted by additive serially correlated noise and that the noise is, in fact, induced by a general linear process. The main result of this study is that, under some reasonable conditions, the nonparametric kernel estimator of $m(x)$ is asymptotically normally distributed. Using this result, we construct confidence bands for $m(x)$. Simulations will be conducted to assess the performance of these bands in finite-sample situations.

AMS 1991 subject classification: 60G70, 62G99, 62G30.
Keywords: Asymptotic normality, correlated errors, nonparametric estimation, time series-analysis.

1 Introduction

Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a strictly stationary bivariate random data sampled either with stochastic design random variables $X_1, X_2, \ldots, X_n$. In the stochastic design model, $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ are identically distributed as a bivariate random variable (X, Y) with values in $\mathbb{R} \times \mathbb{R}$. We suppose that (X, Y) admits a density function $g(x, y)$. Then, we may define the functional parameters

$$f(x) = \int g(x, y)dy, \qquad x \in \mathbb{R},$$

$$\varphi(x) = \int yg(x, y)dy, \qquad x \in \mathbb{R},$$

and

$$m(x) = \varphi(x)/f(x), \qquad \text{if} \quad f(x) > 0. \tag{1}$$

We will say that $m(x)$ is the regression function of Y given $X = x$. The problem is to construct an estimator of m based on the data (X_t, Y_t), $1 \leqslant t \leqslant n$, such that

$$Y_t = m(X_t) + u_t, \qquad t \geqslant 1. \tag{2}$$

Here the noise process $\{u_t\}$ is defined by the stochastic difference equation

$$u_t = \sum_{i=0}^{\infty} \lambda_i \epsilon_{t-i}, \qquad t = 0, \pm 1, \pm 2, \ldots \tag{3}$$

where λ_0=1, $\sum_{i=0}^{\infty} |\lambda_i| < \infty$ and $\{\epsilon_t, t \geqslant 1\}$ is a sequence of independently and identically distributed (*i.i.d.*) random variables such that $E(\epsilon_t)$=0 and $E(\epsilon_t^2) = \sigma^2 < \infty$. This setup covers virtually all commonly used time series models including the well-known autoregressive moving average (ARMA) models.

In classical parametric models, m belongs to a known parametrized class of functions. If this class is not correctly specified, the parametric analyses may lead to misinterpretations of the underlying functional relationship. For complex relationships, nonparametric approaches provide a more flexible alternative. A large class of nonparametric estimators of $m(x)$ is given by

$$m_n(x) = \sum_{t=1}^{n} \omega_{n,t}(x) Y_t, \tag{4}$$

where $\{\omega_{n,t}(x)\}$, $t = 1, \ldots, n$, are weight sequences depending on the distance between the argument x, the vector of regression observations Y_t and on the number of observations. A variety of weight functions exists such as kernel weights, splines, and orthogonal series smoothing; see, e.g., Härdle (1990). For the kernel regression estimator, we use the so-called Nadaraya-Watson estimator. As for the density, the method uses a convolution kernel which regularizes the empirical measures. Let $\delta_{(.)}$ be the Dirac delta function. Consider the empirical measure

$$\lambda_n = \frac{1}{n} \sum_{t=1}^{n} \delta_{(X_t, Y_t)}$$

and its marginal distribution

$$\mu_n = \frac{1}{n} \sum_{t=1}^{n} \delta_{(X_t)}.$$

A regularization of λ_n and μ_n by convolution leads to natural estimators of f_n and φ_n:

$$f_n(x) = \frac{1}{nh_n} \sum_{t=1}^{n} K\left(\frac{x - X_t}{h_n}\right), \qquad x \in \mathbb{R},$$

and

$$\varphi_n(x) = \frac{1}{nh_n} \sum_{t=1}^{n} Y_t K\left(\frac{x - X_t}{h_n}\right), \qquad x \in \mathbb{R},$$

where K is a strictly positive density function and where $\{h_n\}$ is a smoothing parameter which satisfies

$$\lim_{n \to \infty} h_n = 0.$$

Consequently, the kernel estimator of $m(x)$ is defined, for each $x \in \mathbb{R}$, as

$$\begin{aligned} m_n(x) &= \varphi_n(x)/f_n(x) \\ &= \sum_{t=1}^{n} \omega_{n,t}(x) Y_t \end{aligned} \tag{5}$$

where

$$\omega_{n,t}(x) = K\left(\frac{x - X_t}{h_n}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{x - X_j}{h_n}\right).$$

Note that if K is not strictly positive, definition (4) must be completed; i.e. if $f_n(x) = 0$, one may choose

$$m_n(x) = \frac{1}{n} \sum_{t=1}^{n} Y_t$$

which is clearly more natural than the arbitrary $m_n(x) = 0$ used by many authors.

The basic fixed-design stochastic sequence model is not new. In the case of independent noise, it goes back to Priestley and Chao (1972), Gasser and Müller (1979), Michels (1992), Hurvich and Simonoff (1998) and others. When the errors are dependent, the estimation problem generally becomes more difficult to handle. Chu and Marron (1991) have considered the case where the noise process follows an $ARMA$ process. However, they were concerned with bandwidth estimation, not convergence of the regression estimate. Truong (1991) obtained optimal rates of kernel estimators based on local averages. Härdle and Tuan (1986) and Truong (1991) studied the case where the noise process is a general linear process. When the u_t's satisfy the strong mixing (or α-mixing) conditions, a suitable normalized version of $m_n(x)$ has been shown to be asymptotically normal by Roussas, Tran and Ioannides (1992). Further, Burman (1991) constructed a spline estimate of the regression function and obtained its rate of convergence. Also, in the case of fixed-design regression and serially correlated noise process, Tran, Roussas, Yakowitz and Truong Van (1996) showed the asymptotic normality of $m_n(x) - E(m_n(x))$. However, within applied econometrics and time series analysis models with one or more random input series X_t are commonly used. Examples include transfer function noise models and dynamic regression models. In the random-design case, Truong and Stone (1992) and Tran (1993) obtained optimal rates of convergence of local average estimators under α-mixing conditions. Roussas and Tran (1992) have established asymptotic normality of recursive regression estimators under related conditions.

In this paper, we investigate the asymptotic behaviour of the estimator $m_n(x)$ in the case $\{X_t\}$ is a strictly stationary discrete-time stochastic process and the noise process $\{u_t\}$ are serially correlated. The main result of this study is that, under some reasonable conditions, $m_n(x)$ is asymptotically normally distributed. Using this result, we construct a confidence band for $m(x)$. Simulations will be conducted to assess the performance of these bands in finite-sample situations. Our results are applicable to a variety of nonparametric functions classes under different correlation structures of the noise process.

The organization of this paper is as follows: in Section 2 we present the assumptions and conditions under which the main theorem is true, followed by the statement of the main result. The useful and relevant technical lemmas needed for proving the main result are given in Section 3. Section 4 contains simulations results for two confidence bands for $m(x)$ for nonparametric regression with a random X_t and serially correlated errors. Finally, two empirical examples are given in Section 5.

2 Assumptions and statement of main result

The assumptions under which the result in this paper is derived are gathered together below for easy reference. It will be usually indicated which of these assumptions are used in the various steps of derivations.

Assumptions on the stochastic process:

(H.1) $E(Y_t^2|X_t = .) < \infty$; $|E(Y_t^3|X_t = .)| < \infty$, $\forall t \geqslant 1$;

(H.2) The function m is bounded. Further the functions m and f are twice differentiable, and the partial derivatives are bounded and continuous.

(H.3) The variables X_t and u_t are independent.

(H.4) $\sum_{i=0}^{\infty} |\lambda_i| < \infty$, $\sum_{i=0}^{\infty} \lambda_i^2 < \infty$.

(H.5) $\{\epsilon_t, t \geqslant 1\}$ is a sequence of *i.i.d.* random variables such that $E(\epsilon_t)$=0 and $E(\epsilon_t^2) = \sigma^2 < \infty$.

Assumption on the kernel:

(K.1) The function K is bounded such that $y^2 K(y) \to 0$ when $y \to \infty$, $\int_{-\infty}^{\infty} y^2 K(y) dy < \infty$, and $\int_{-\infty}^{\infty} y^3 K(y) dy < \infty$

Assumption on the bandwidth:

(M.1) $nh_n \to \infty$ and $nh_n^5 \to 0$ when $n \to \infty$.

We now proceed with the formulation of the main result of the paper.

Theorem 1. *Let $x \in I\!R$. Suppose assumptions (H.1)–(H.5), (K.1), and (M.1) hold. Then*

$$\frac{\sqrt{nh_n}\{m_n(x) - m(x)\}}{\sqrt{\sigma^2 (\sum_{i=0}^{\infty} \lambda_i^2) \frac{1}{f(x)} \int K^2(y) dy}} \xrightarrow{d} N(0,1). \qquad (6)$$

Let us mention an important consequence of the above theorem.

Corollary 1. *Let $x \in C$ a compact set in $I\!R$. Suppose that assumptions (H.1)–(H.5), (K.1), and (M.1) hold. In addition we assume that the following assumptions are satisfied (C.1) K satisfies the Lipschitz condition; (C.2) $nh_n/Log\, n \to \infty$; and (C.3) the ϵ_t's are i.i.d. $N(0,1)$ distributed. Then*

$$\frac{\sqrt{nh_n}\{m_n(x) - m(x)\}}{\sqrt{s_n^2 \frac{1}{f_n(x)} \int K^2(y)dy}} \xrightarrow{d} N(0,1), \qquad (7)$$

where $s_n^2 = n^{-1} \sum_{t=1}^n (Y_t - m_n(X_t))^2$.

3 Some intermediate results

In this section, some additional auxiliary results are derived, necessary to establish the asymptotic normality of m_n. First we introduce some new notations. For $t = 1, \dots, n$, we define

$$\begin{aligned}
U^*_{n,t}(x) &= \frac{1}{h_n} K(\frac{x - X_t}{h_n}), \qquad U_{n,t}(x) = \sqrt{h_n}\ \{U^*_{n,t}(x) - E(U^*_{n,t}(x))\}, \\
V^*_{n,t}(x) &= \frac{1}{h_n} K(\frac{x - X_t}{h_n}) Y_t, \qquad V_{n,t}(x) = \sqrt{h_n}\{V^*_{n,t}(x) - E(V^*_{n,t}(x))\}, \\
U_n(x) &= \sum_{t=1}^n U_{n,t}(x), \qquad V_n(x) = \sum_{t=1}^n V_{n,t}(x), \\
W_{n,t}(x) &= \begin{pmatrix} U_{n,t}(x) \\ V_{n,t}(x) \end{pmatrix}, \qquad \sqrt{n} Z_n(x) = \begin{pmatrix} U_n(x) \\ V_n(x) \end{pmatrix},
\end{aligned}$$

and

$$\sqrt{n} Z^*_n(x) = \sqrt{h_n} \begin{pmatrix} \sum_{t=1}^n \{U^*_{n,t}(x) - f(x)\} \\ \sum_{t=1}^n \{V^*_{n,t}(x) - \varphi(x)\} \end{pmatrix}.$$

Now, let

$$A(x) = f(x) \int K^2(y)dy \begin{pmatrix} 1 & m(x) \\ m(x) & E(Y_t^2 | X_t = x) \end{pmatrix}$$

be a positive definite matrix.

Next we have to study the asymptotic normality of $Z_n(x)$, and to compute its asymptotic variance-covariance matrix. So, we need the following lemma.

Lemma 1. *Under (H.2), (K.1) and (M.1), for every continuity point x of f, $x \in C(f)$,*

(a) $\lim_{n\to\infty} E(U_{n,t}^2(x)) = f(x) \int K^2(y)dy$*;*

(b) $\lim_{n\to\infty} E(V_{n,t}^2(x)) = f(x) E(Y_t^2|X_t = x) \int K^2(y)dy$*;*

(c) $\lim_{n\to\infty} E(U_{n,t}(x)V_{n,t}(x)) = f(x)m(x) \int K^2(y)dy = \varphi(x) \int K^2(y)dy$*.*

Proof. The results (a), (b), and (c) follow directly using Bochner's lemma; see, e.g., Roussas (1990, Proposition 2.1). □

The next result concerns the asymptotic normality of $Z_n(x)$.

Lemma 2. *Let $x \in I\!R$. Under (H.1)–(H.5), (K.1) and (M.1)*

$$Z_n(x) \xrightarrow{d} N(0, A(x)),$$

where $N(0, A(x))$ denotes a bivariate normally distributed random variable with mean 0 and variance $A(x)$.

Proof. The basic idea of the proof is to establish the conditions for the application of the version of the Central Limit Theorem given by Liapounov (see, e.g., Loève 1955, p. 275), and the Cramér-Wold theorem (see, e.g., Rao, 1973). Let $c = (c_1, c_2)'$ be a 2×1 column vector where $(c_1, c_2) \in \mathbb{R}^2$ such that $c_1^2 + c_2^2 \neq 0$. For $t = 1, \ldots, n$, let $\sigma_{n,t}^2(x) = Var(c'W_{n,t}(x))$ and $\rho_{n,t}^3(x) = E(|c'W_{n,t}(x)|^3)$. Assume $\sigma_n^2(x) = \sum_{t=1}^n \sigma_{n,t}^2(x)$ and $\rho_n^3(x) = \sum_{t=1}^n \rho_{n,t}^3(x)$. Using Lemma 1, we have

$$\begin{aligned} \lim_{n\to\infty} \sigma_{n,t}^2(x) &= c_1^2 f(x) \int K^2(y)dy + c_2^2 f(x) E(Y_t^2|X_t = x) \int K^2(y)dy \\ &\quad + 2c_1c_2 f(x)m(x) \int K^2(y)dy \\ &= c'A(x)c. \end{aligned} \tag{8}$$

Now, using Bochner's lemma and the assumptions (H.1) and (K.1) we get

$$E(|U_{n,1}(x)|^3) \sim h_n^{-1/2} f(x) \int K^3(y)dy$$

and hence

$$E(|U_{n,1}(x)|^3) = O(h_n^{-1/2}).$$

Similarly, we have

$$E(|V_{n,1}(x)|^3) \sim h_n^{-1/2} f(x) E(Y_t^3 | X_t = x) \int K^3(y) dy.$$

Thus,

$$E(|V_{n,1}(x)|^3) = O(h_n^{-1/2}).$$

Consider now $\rho_n^3(x)$:

$$\begin{aligned} \rho_n^3(x) &= nE(|c' W_{n,1}(x)|^3) \\ &\leqslant |c_1|^3 nE(|U_{n,1}|^3) + |c_2|^3 nE(|V_{n,1}|^3) \\ &\quad + 3|c_1|c_2^2 nE(|U_{n,1}V_{n,1}^2|) + 3c_1^2|c_2|nE(|U_{n,1}^2 V_{n,1}|). \end{aligned}$$

The above two results and some elementary calculations give $\rho_n^3(x) = O(nh_n^{-1/2})$. Expression (8) implies that $\sigma_n^3(x) = O(n^{3/2})$. Thus, using assumption (M.1), we obtain

$$\lim_{n\to\infty} \frac{\rho_n^3(x)}{\sigma_n^3(x)} = \lim_{n\to\infty} \frac{1}{\sqrt{nh_n}} = 0.$$

Thus, we established the conditions for the application of the version of the Central Limit Theorem given by Liapounov. Then we can say that

$$c' Z_n(x) = n^{-1/2} \sum_{t=1}^{n} c' W_{n,t}(x)$$

converges in law to a normal distribution with mean 0 and variance $c'A(x)c$. The proof of Lemma 2 is completed by using the Cramér-Wold theorem. □

Now, we have to prove that $Z_n^*(x)$ is asymptotically normally distributed. We begin by proving the following lemma.

Lemma 3. *Let assumptions (H.2), (K.1) and (M.1) hold. Then for $(x, y) \in I\!R^2$, we have*

$$\lim_{n\to\infty} (nh_n)^{1/2} \{E(U_{n1}^*(x)) - f(x)\} = 0,$$
$$\lim_{n\to\infty} (nh_n)^{1/2} \{E(V_{n1}^*(x)) - \varphi(x)\} = 0.$$

Proof. By direct computation, it can be easily shown that

$$(nh_n)^{1/2}\{E(U_{n1}^*(x)) - f(x)\} = O(nh_n^5)^{1/2}$$

$$(nh_n)^{1/2}\{E(V_{n1}^*(x)) - \varphi(x)\} = O(nh_n^5)^{1/2}.$$

The proof is completed by using (M.1). □

According to the above result and Lemma 2, we have the following lemma.

Lemma 4. *Suppose assumptions (H.1)–(H.5), (K.1) and (M.1) are satisfied. Then*

$$Z_n^*(x) \xrightarrow{d} N(0, A(x)).$$

Proof.

$$\begin{aligned} Z_n^*(x) &= Z_n(x) + (Z_n^*(x) - Z_n(x)) \\ &= Z_n(x) + \sqrt{nh_n} \begin{pmatrix} E(U_{n,1}^*(x)) - f(x) \\ E(V_{n,1}^*(x)) - \varphi(x) \end{pmatrix}. \end{aligned}$$

The proof is completed by using Lemmas 2 and 3. □

We now prove Theorem 1 of this paper.

Proof of Theorem 1. The proof uses the Mann-Wald Theorem (see Rao (1973), p. 388). Let the function $H : \mathbb{R}^2 \to \mathbb{R}$ defined by $H(y_1, y_2) = y_2/y_1$, with $y_1 \neq 0$, and let

$$\theta(x) = \begin{pmatrix} f(x) \\ \varphi(x) \end{pmatrix} \quad \text{and} \quad T_n(x) = \begin{pmatrix} f_n(x) \\ \varphi_n(x) \end{pmatrix}.$$

Using the notation introduced at the beginning of this section:

$$T_n(x) = \begin{pmatrix} \frac{1}{n}\sum_{t=1}^n U_{n,t}^*(x) \\ \frac{1}{n}\sum_{t=1}^n V_{n,t}^*(x) \end{pmatrix}.$$

Now, we can write

$$Z_n^*(x) \;=\; (nh_n)^{1/2}(T_n(x) - \theta(x)).$$

Hence

$$(nh_n)^{1/2}\{m_n(x) - m(x)\} = (nh_n)^{1/2}\{H(T_n(x)) - H(\theta(x))\}.$$

Using Lemma 4, we have $\sqrt{nh_n}(T_n(x) - \theta(x)) \xrightarrow{d} N(0, A(x))$ and applying Mann-Wald's theorem (or the δ-method), it follows that $(nh_n)^{1/2}\{m_n(x) - m(x)\}$ converges in law to the normal distribution with mean 0 and variance $V(x)$ defined by

$$V(x) = D'(x)A(x)D(x),$$

where $D(x)$ be the column vector of partial derivatives of H with respect to $\theta(x)$, i.e.

$$\begin{aligned} D(x) &= \Big(-\frac{\varphi(x)}{f^2(x)} \quad \frac{1}{f(x)} \Big)' \\ &= \frac{1}{f(x)}\Big(-m(x) \quad 1 \Big)'. \end{aligned}$$

Now, $V(x)$ is given by

$$V(x) = \Big\{ E(Y_t^2|X_t = x) - m^2(x) \Big\} \frac{1}{f(x)} \int K^2(y)dy.$$

Furthermore, we have

$$E(Y_t^2|X_t = x) - m^2(x) = E\{(m(X_t) + u_t)^2|X_t = x\} - m^2(x).$$

Since the variables X_t and u_t are independent and $E(u_t) = 0$ $\forall t$, we have

$$E(Y_t^2|X_t = x) - m^2(x) = E(u_t^2).$$

Then, with $E(u_t^2) = \sigma^2(\sum_{i=0}^{\infty} \lambda_i^2)$, we have

$$V(x) = \sigma^2(\sum_{i=0}^{\infty} \lambda_i^2) \frac{1}{f(x)} \int K^2(y)dy$$

which completes the proof. □

Proof of Corollary 1. First, we have to prove that $s_n^2 \xrightarrow{p} \sigma^2 \sum_{i=0}^{\infty} \lambda_i^2$. Obviously, s_n^2 can be written as follows

$$\begin{aligned} s_n^2 &= n^{-1}\sum_{t=1}^{n}(Y_t - m(X_t))^2 + n^{-1}\sum_{t=1}^{n}(m(X_t) - m_n(X_t)))^2 \\ &\quad + 2n^{-1}\sum_{t=1}^{n}(Y_t - m(X_t))(m(X_t) - m_n(X_t)). \end{aligned} \tag{9}$$

Using Proposition 7.3.5 of Brockwell and Davis (1996) we get for the first term on the right-hand side of (9)

$$n^{-1}\sum_{t=1}^{n}(Y_t - m(X_t))^2 = n^{-1}\sum_{t=1}^{n} u_t^2 \xrightarrow{p} \sigma^2 \sum_{i=0}^{n} \lambda_i^2.$$

Now, we show that the remaining two terms converge to 0 in probability. Clearly, the second term of (9) can be written as

$$n^{-1}\sum_{t=1}^{n}(m(X_t) - m_n(X_t))^2 = n^{-1}\sum_{t=1}^{n}(m(X_t) - E(m_n(X_t))^2$$

$$+n^{-1}\sum_{t=1}^{n}(m_n(X_t) - E(m_n(X_t))^2 + 2n^{-1}\sum_{t=1}^{n}(m(X_t) - E(m_n(X_t))(m_n(X_t) - E(m_n(X_t)).$$

Using Bosq (1996, Theorem III.1) it follows that

$$\sup_{x\in C}\Big(m(x) - E(m_n(x))\Big)^2 = \sup_{x\in C}\Big(\Big[\frac{h_n^2}{2}\xi(x) + o(h_n^2) + o(\frac{1}{nh_n})^2\Big]\Big)^2$$

where

$$\xi(x) = \Big(\frac{d^2m}{dx^2} + \frac{d\mathrm{Log}\, f}{dx}\frac{dm}{dx}\Big)(x)\int t^2K(t)dt$$

and

$$\sup_{x\in C}(m_n(x) - E(m_n(x))^2 = \sup_{x\in C}\Big[\frac{1}{nh_n}\frac{v(x)}{f(x)}\int K^2(y)dy + o\Big(h_n^4 + \frac{1}{nh_n}\Big)\Big]$$

where $v(x) = V(Y|X = x)$ that is bounded using assumptions (H.2), (H.4) and (H.5). So, under assumption (M.1), the term $n^{-1}\sum_{t=1}^{n}(m(X_t) - E(m_n(X_t))^2$ converges to 0 and $n^{-1}\sum_{t=1}^{n}(m_n(X_t) - E(m_n(X_t)))^2$ converges to 0 in probability.

Now

$$\begin{aligned} &\Big|n^{-1}\sum_{t=1}^{n}(m(X_t) - E(m_n(X_t)))(m_n(X_t) - E(m_n(X_t)))\Big| \leqslant \\ &\qquad \sup_{x\in C}\Big|m(x) - E(m_n(x))\Big| \sup_{x\in C}\Big|m_n(x) - E(m_n(x))\Big|. \end{aligned} \tag{10}$$

Using Bernstein-Fréchet inequality, under assumption (C.2), the second term on the right-hand side of (10) tends to 0, i.e.,

$$\sup_{x\in C}\Big|m_n(x) - E(m_n(x))\Big| \xrightarrow[n\to\infty]{p} 0.$$

So the second term of (9) converges to 0 in probability. Now we consider the third term on the right-hand side of (9)

$$\begin{aligned} n^{-1}\sum_{t=1}^{n}(Y_t - m(X_t))(m(X_t) - m_n(X_t)) &= n^{-1}\sum_{t=1}^{n} u_t(m(X_t) - m_n(X_t)) \\ &\leqslant \sup_{x\in C}\Big|m(x) - m_n(x)\Big|\ n^{-1}\sum_{t=1}^{n} u_t. \end{aligned}$$

If $\gamma(.)$ is the autocovariance function of $\{u_t\}$, then for any non-negative integer k, $|\gamma(k)| = |\sum_{i=0}^{\infty}\lambda_i\lambda_{i+k}| \leqslant \sqrt{\sum_{i=0}^{\infty}\lambda_i^2}\sqrt{\sum_{i=0}^{\infty}\lambda_{i+k}^2}$. Under assumption (H.4), $\sum_{i=0}^{\infty}\lambda_{i+k}^2 \xrightarrow{k\to\infty} 0$. So $|\gamma(k)| \xrightarrow{k\to\infty} 0$. We now apply Theorem 5 of Gihman-Skorokhod (1980, p. 158). The process $\{u_t\}$ is ergodic, so $n^{-1}\sum_{t=1}^{n} u_t \xrightarrow{ps} E(u_t) = 0$. Moreover, $\sup_{x\in C}|m(x) - m_n(x)| \xrightarrow{p} 0$. So, $n^{-1}\sum_{t=1}^{n}(Y_t - m(X_t))(m(X_t) - m_n(X_t)) \xrightarrow{p} 0$.

Hence, we have

$$s_n^2 \xrightarrow{p} \sigma^2\sum_{i=0}^{\infty}\lambda_i^2. \tag{11}$$

Now it is clear that

$$\frac{\sqrt{nh_n}\{m_n(x) - m(x)\}}{\sqrt{s_n^2\frac{1}{f_n(x)}\int K^2(y)dy}} = \underbrace{\frac{\sqrt{\sigma^2(\sum_{i=0}^{\infty}\lambda_i^2)\frac{1}{f(x)}\int K^2(y)dy}}{\sqrt{s_n^2\frac{1}{f_n(x)}\int K^2(y)dy}}}_{(A)} \times \underbrace{\frac{\sqrt{nh_n}\{m_n(x) - m(x)\}}{\sqrt{\sigma^2(\sum_{i=0}^{\infty}\lambda_i^2)\frac{1}{f(x)}\int K^2(y)dy}}}_{(B)}.$$

According to Theorem 1, term (B) is asymptotically normally distributed. From (11), term (A) converges to 1 in probability because $1/f_n(x)$ converges in probability to $1/f(x)$; see, e.g., Bosq and Lecoutre (1987). The proof of Corollary 1 is completed by using Slutzky's theorem. $\square$

4 Simulations

To assess the performance of the confidence bands that can be constructed from Theorem 1 and Corollary 1 in Section 2, we conduct a Monte Carlo experiment. Two sets of regression functions will be considered:

I) $m(X_t) = 50X_t^3(1 - X_t)^3$;

II) $m(X_t) = 2.3\exp\{-\frac{1}{2}X_t^2\}$.

Function I) is taken from Chu and Marron (1991) and Tran et al. (1996). The regressor X_t is sampled from the AR(1) process $X_t = 0.5X_{t-1} + \xi_t$ with $\xi_t \overset{i.i.d.}{\sim} N(0,1)$. For $\{u_t\}$ we choose the following two processes:

a) AR(1) process: $u_t = \phi u_{t-1} + \epsilon_t$ with $\epsilon_t \overset{i.i.d.}{\sim} N(0,1)$ and $|\phi| < 1$;

b) MA(1) process; $u_t = \epsilon_t + \theta\epsilon_{t-1}$ with $\epsilon_t \overset{i.i.d.}{\sim} N(0,1)$ and $|\theta| < 1$.

Simulation results will be reported for samples sizes n=100, 200, 300, 500, and 1000. For a single replicate sample of the regression function $m_n(X_t)$, r=100 replicates of the noise process $\{u_t\}$ are generated. Then we obtain the estimate of the nonparametric regression of n equally spaced values $X_{(1)}, X_{(2)}, \ldots, X_{(n)}$ in the $(0, 1]$ interval, where $X_{(i)} = \min(X_t) + (i-1) \times (range(X_t)/n)$ with $(i, t = 1, \ldots, n)$. Next, the whole experiment is repeated until altogether N=100 replicates samples of the regressor $m_n(X_t)$ are generated. As a measure to evaluate $m_n(x)$ we compute for each replicate of the regression function the Averaged Squared Error (ASE)

$$ASE(N) = \frac{1}{100n}\sum_{r=1}^{100}\sum_{i=1}^{n}\{m_n^r(X_{(i)}) - m(X_{(i)})\}^2, \quad (N = 1, \ldots, 100),$$

where r is the replicate number of the noise process. In addition the overall average of the $ASE(N)$'s is computed, i.e. $ASE = 100^{-1}\sum_{N=1}^{100} ASE(N)$.

To assess the performance of the confidence bands defined by Corollary 1, we record the proportion of times the true regression values $m(X_t)$ are contained in the confidence bands

$$m_n(X_t) \pm z_\alpha s_n \sqrt{\int K^2(y)dy / \sqrt{nh_n f_n(x)}} \tag{12}$$

where z_α is the $100(1-\alpha)$th percentile of the standard normal distribution. For α=0.05 and α=0.1, we also compute the average (over the N=100 replicates) of these empirical coverage probabilities. For the estimation of $m(x)$ the Gaussian kernel function is used throughout the calculations which implies that $\int K^2(y)dy = 1/2\sqrt{\pi}$ in Corollary 1. The bandwidths are chosen by using $h_n = 1.06n^{-1/5}\hat{\sigma}_n$ where $\hat{\sigma}_n$ is the estimated standard deviation of the series X_t.

Table 1 about here

Table 1 shows the empirical coverage probabilities for the particular case $\{u_t\}$ follows a white noise (WN) error process, an AR(1) process with parameter ϕ=0.5, and an MA(1) process with parameter θ=0.5. It is evident that as n increases, the empirical coverages approach the nominal probabilities. For the WN case, we can see that at n=1000 the empirical coverages are nearly nominal coverages for both regression functions. However, for the AR(1) case with n=1000 the results are still far from desired. On the other hand, the empirical coverage probabilities in the MA(1) case improve very gradually with sample sizes increases.

Figure 1 about here

Examination of Figure 1.a clarifies the differences between the nominal and empirical coverage probabilities in the AR(1) and MA(1) case. There, for model I), α=0.05, and n=300, 500, and 1000, we plotted the curves of the empirical coverage probabilities versus the AR(1) error process parameters $\phi = -0.95, -0.90, \ldots, 0.95$. Figure 1.a also contains a conservative 95% critical region about the nominal confidence level, indicated by two dash-dotted (practically horizontal) lines, and computed as $0.95 \pm 2 \times SD(cov)$, where $SD(cov)$ is the standard deviation of the overall (averaged over N=100 replicates and for n=1000) empirical coverage probabilities at each AR parameter value ϕ. This critical region provides a quick means of determining whether any of the empirical coverages are significantly different from the nominal level at the 5% level. As can be noted from Figure 1.a the difference between the nominal and empirical coverage probabilities is not greater than 5% if the values of the AR parameters ϕ are within the range [–0.75,0.0]. Thus if we approximate the MA(1) error process in Table 1 by an AR(1) error process with ϕ=–0.5,

Table 1: Empirical coverage probabilities and overall ASE (multiplied by 10^4) for the particular case $\{u_t\}$ follows a white noise error process, an AR(1) process with parameter ϕ=0.5, and an MA(1) process with parameter θ=0.5, for models I) and II), and for various sample sizes n.

Noise process	n	Regression function I)			Regression function II)		
		Nominal coverage (%)			Nominal coverage (%)		
		95	90	ASE	95	90	ASE
WN	100	0.924	0.864	326.121 (28.208)	0.952	0.904	271.651 (23.266)
	200	0.929	0.870	180.705 (12.286)	0.956	0.910	150.052 (10.927)
	300	0.930	0.871	128.512 (9.102)	0.956	0.911	106.556 (7.238)
	500	0.932	0.874	83.903 (5.950)	0.959	0.914	69.437 (4.331)
	1000	0.932	0.874	46.519 (2.777)	0.961	0.917	36.754 (2.187)
AR(1)	100	0.840	0.761	657.224 (49.053)	0.862	0.786	602.797 (44.351)
	200	0.862	0.787	340.593 (24.430)	0.885	0.812	310.099 (23.609)
	300	0.867	0.793	236.688 (17.035)	0.889	0.819	215.034 (15.746)
	500	0.881	0.808	147.104 (9.294)	0.904	0.835	132.773 (7.909)
	1000	0.890	0.820	78.824 (4.075)	0.915	0.850	68.904 (3.757)
MA(1)	100	0.889	0.819	489.721 (38.332)	0.914	0.849	435.254 (34.356)
	200	0.900	0.834	261.243 (17.583)	0.925	0.865	230.765 (16.608)
	300	0.904	0.837	183.464 (12.711)	0.928	0.869	161.719 (11.093)
	500	0.912	0.847	116.931 (7.895)	0.935	0.878	102.535 (6.334)
	1000	0.916	0.853	63.504 (3.604)	0.941	0.887	53.691 (3.149)

the empirical coverage probability will lie within the above critical region. As expected if ϕ approaches the stationary boundary, the discrepancy between the nominal and empirical coverage probabilities increases. At ϕ=0.95 the empirical coverage probabilities are as low as 0.773, 0.799, and 0.829 for respectively n=300, 500 and 1000. The effect of sample size on the differences between the nominal and empirical coverage probabilities becomes more apparent as $\phi \to 1$. If n=300 the potential difference is about 18.6% whereas for n=1000 it is only 12.7%. Both these percentages may be a cause for some concern. Nevertheless evidence of much larger differences ($> 20\%$) have been reported in many practical situations; see, e.g., Donaldson and Schnabel (1987).

Table 1 also contains values of the ASE with estimates of the standard deviation of the ASE in parentheses. We can see that the ASE decreases with increased sample size. Figure 1.b shows curves of the ASE versus the AR(1) parameters $\phi = -0.95, -0.90, \ldots, 0.95$ for model I) with n=300, 500 and 1000. The general features of these curves are essentially the same. However, notice that for n=300 the ASE is about 3 times larger than for n=1000. Hence, there is a significant loss of efficiency in the estimation of $m(x)$ by $m_n(x)$ if the sample size is too small. Note that the gaps between the three curves widen as $\phi \to 1$. This shows that the loss of estimation efficiency even becomes higher for small n when the errors are strongly positively serially correlated. It means that the rate of convergence of the estimator is dominated by the near-nonstationarity of the error process.

5 Examples

As a first example, consider the systolic (X_t) and diastolic (Y_t) blood pressure readings given in Figure 2.a. The data were sampled twice a day for 230 days (n=460); see Shumway (1988, Appendix I, Table 5). The interest was to determine a possible leading-lagging linear relationship between the two series. The sample autocorrelation function (ACF) of both series showed a slow positive decay from a value of 0.42 (0.37) at lag 1 to a value of 0.25 (0.22) at lag 10 for X_t (Y_t). These values are all significant at the 5% level, suggesting clearly that first differencing of the series is necessary. On doing this, the sample ACFs of the series $\Delta X_t = X_t - X_{t-1}$ and $\Delta Y_t = Y_t - Y_{t-1}$ are almost compatible with white noise, except from significant negative spikes at lag 1. The sample cross correlation function (CCF) of ΔY_t and ΔX_t showed a significant value of 0.35 at lag 0. Also small (-0.18 and -0.13), though significant, values of the sample CCF were obtained at lags 1 and -1, respectively. This provides some information about how ΔY_t is related to current and earlier values of ΔX_t.

Figure 2 about here

Figure 2.b shows a plot of the smoothed estimate $m_n(\Delta X_t)$ (full curve) from applying (5) (with h_n=2.17) to the series ΔY_t and ΔX_t, surrounded by the 95% confidence bands (medium dashed lines) obtained from Corollary 1. It is apparent from this figure that the width of these bands increases markedly as the values of ΔX_t approach the left and right boundaries of the observation interval. This is due to the fact that the data are more sparse near these boundaries. From the central part of Figure 2.b we see that there exists a nearly linear relationship between ΔY_t and ΔX_t with a positive slope parameter.

We checked the adequacy of the nonparametric regression by examining the residual ACF of the series $\Delta Y_t - m_n(\Delta X_t)$. All but the residual ACF at lag 1 (–0.48) lie well within the 95% confidence bands. This may indicate that the residuals themselves follow an MA(1) process. To check for any leading-lagging relationship between ΔY_t and ΔX_t we estimated a dynamic regression (DR) model with ΔY_t related to its own past and the past of ΔX_t up to lag 2, and with a first-order MA error term. All estimated coefficients appeared to be not significantly different from zero at the 5% level, except for the coefficient of ΔX_t and the MA(1) parameter. The "final" model (with approximate standard errors in parentheses) fitted to the data is given by

$$\begin{array}{rcllll} \Delta Y_t & = & -0.009+ & 0.284\Delta X_t + \epsilon_t & -0.860\epsilon_{t-1} \\ & & (0.022) & (0.028) & (0.023) \end{array} \qquad (13)$$

with R^2=0.48, the residual standard deviation is $\hat{\sigma}$=3.35, and the value of the Durbin-Watson (DW) statistic is 1.90. The residual ACF indicated that the errors are white noise, as they should be for a correct fit. The fitted regression line is plotted in Figure 2.b (dotted line). Note that model (13) is different from the best "fitted" model $Y_t = 0.6X_t$ proposed by Shumway (1988, p. 104). A reasonable interpretation of (13) is that a change in the systolic blood pressure leads to an almost instantaneous change in the diastolic blood pressure.

The second example concerns the declining number of ovarian follicles (eggs) as women age. The data set contains the age of 110 females in years, and the count of follicles; see the website: www.blackwellpublishers.co.uk/rss/. Faddy and Jones (1999) analysed the data by local quadratic nonparametric smoothing. Their interest was in the pattern of decline in the number of follicles as a function of age, leading to infertility and the menopause. Hence, the input series (X_t) is the age of the women in years. The output series is the common logs of the count of follicles. Figure 3 is a plot of the estimate $m_n(X_t)$ (full curve) together with 95% confidence bands (medium dashed lines) with h_n=5.84. The pattern of decline is nearly linear until age reaches the late 30s. Beyond the age of 40 there appears to be a accelerated loss in fertility. The residual ACF at lags 1 (0.23), 5 (0.21), and 6 (0.25) were all significant at the 5% level. This suggests the presence of first-order serial correlation in the residuals with a parameter value close to 0.23.

Figure 3 about here

A simple way to improve the fitted regression function is to remove the observed AR(1) effect by transforming the data so that $\tilde{Y}_t = Y_t - 0.23Y_{t-1}$. Next, reestimating $m_n(X_t)$ in $\tilde{Y}_t = m_n(X_t) + \epsilon_t$ and checking the residual ACF we noted that the serial correlation was effectively removed. Of course, some caution is needed here because the transformation chosen may not be optimal. Therefore we fitted various parametric DR models to the series. The "best" model, in terms of the minimum value of Akaike's information criterion, is given by

$$\begin{array}{rcllll} Y_t & = & 4.592+ & 0.261Y_{t-1} & -0.040X_t + \epsilon_t \\ & & (0.604) & (0.095) & (0.006) \end{array} \tag{14}$$

with R^2=0.75, $\hat{\sigma}$=0.44, and DW=2.05. The residual ACF indicated that, except at lag 6, no serial correlation was left in the residuals. Clearly, the AR(1) parameter in (14) is very close to the transformation parameter used in the modified nonparametric regression.

References

Bosq, D. and Lecoutre, J.P. (1987). *Théorie de l'estimation fonctionnelle*, Paris: Economica.

Bosq, D. (1996). *Nonparametric Statistics for Stochastic Processes*, Berlin: Springer-Verlag.

Brockwell, P.J. and Davis, R.A. (1996). *Time Series: Theory and Methods*, Berlin: Springer-Verlag (2nd ed.).

Burman, P. (1991). Regression function estimation from dependent observations. *J. Multivariate Anal.* 36, 263-279.

Chu, C. K. and Marron S. (1991). Comparison of two bandwidth selectors with dependent errors. *Annal. Statist.* 19, 1906-1918.

Donaldson, J.R. and Schnabel, R.B. (1987). Computational experience with confidence regions and confidence intervals for nonlinear least squares. *Technometrics* 29, 67-82.

Faddy, M.J. and Jones, M.C. (1999). Modelling and analysis of data that exhibit temporal decay. *Appl. Statistics* 48, 229-237.

Gasser, T. and Müller H.G. (1979). Kernel estimation of regression function. In *Smoothing Techniques for Curve Estimation. Lecture Notes in Math.* 747, 23-68. Berlin: Springer-Verlag.

Gihman, I.I. and Skorokhod, A.V. (1980). *Introduction to the Theory of Random Processes.* Moscow.

Härdle, W. (1990). *Applied Nonparametric Regression*, Econometric Society Monograph No. 19, Cambridge: Cambridge University Press.

Härdle, W. and Tuan, D. P. (1986). Some theory on *M*-smoothing of time series. *J. Time Ser. Anal.* 19, 191-204.

Hurvich, C.M. and Simonoff J.S. (1998). Applied Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. R. Statist. Soc.* B 60, 271-293.

Loève, M. (1955). *Probability Theory*, Princeton: D. Van Nostrand.

Michels, P. (1992). Asymmetric kernel functions in nonparametric regression analysis and prediction, *The Statistician* 41, 439-454.

Priestley, M.B. and Chao, M.T. (1972). Nonparametric function fitting. *J. R. Statist. Soc.* B 34, 385-392.

Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, New York: Wiley (2nd. ed.).

Roussas, G. (1990). Nonparametric regression estimation under mixing conditions. *Stochastic Proceses and its Applications* 36, 107-116.

Roussas, G.G. and Tran, L.T. (1992). Asymptotic normality of the recursive kernel regression estimate under dependence conditions. *Annal. Statist.* 20, 98-120.

Roussas, G.G., Tran, L.T. and Ioannides, D.A. (1992). Fixed design regression for time series: asymptotic normality. *J. Multivariate Anal.* 40, 262-291.

Shumway, R.H. (1988). *Applied Statistical Time Series Analysis*, Englewood Cliffs: Prentice Hall.

Tran, L.T. (1993). Nonparametric function estimation for time series by local average estimators. *Annal. Statist.* 21, 1040-1057.

Tran, L.T., Roussas G.G., Yakowitz, S. and Truong Van B. (1996). Fixed-design regression for linear time series. *Annal. Statist.* 24, 975-991.

Truong, Y.K. (1991). Nonparametric curve estimation with time series errors. *J. Statist. Plann. Inference* 28, 167-183.

Truong, Y.K. and Stone, C.J. (1992). Nonparametric function estimation involving time series. *Annal. Statist.* 20, 77-97.

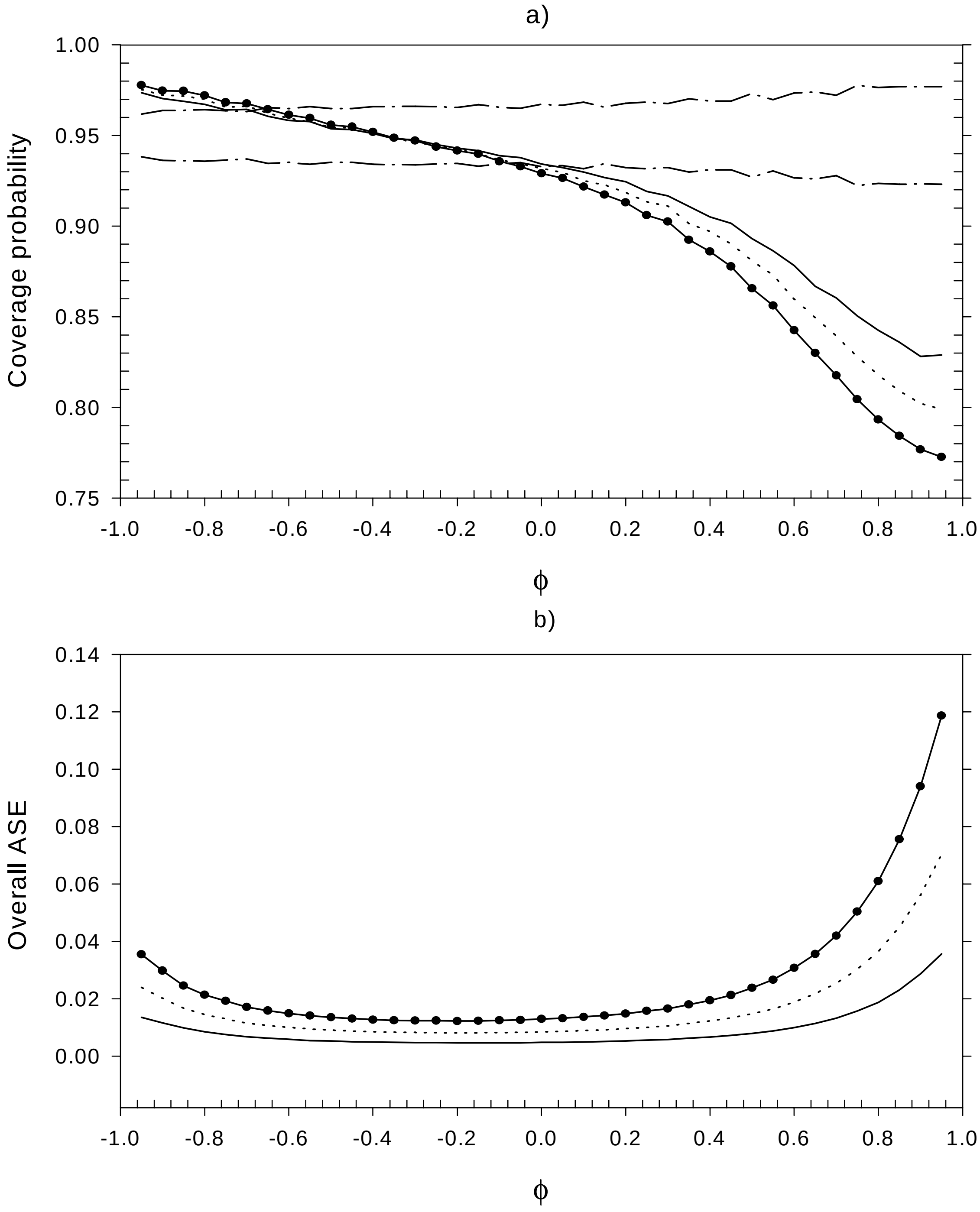
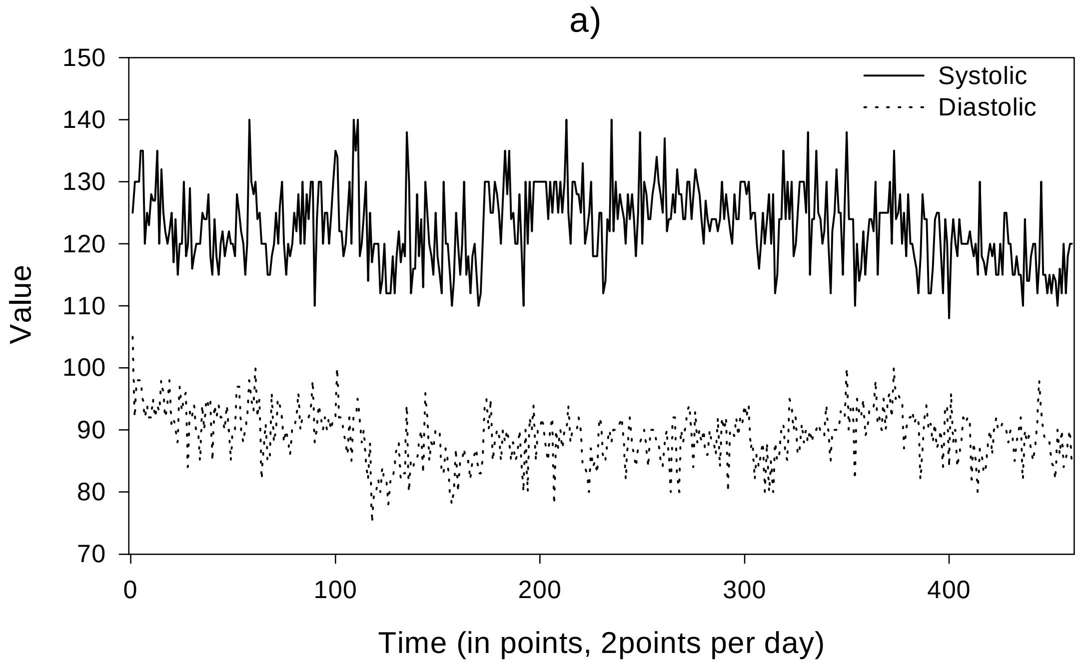


Figure 1: *a) Empirical coverage probabilities for model I) and* α*=0.05 versus the AR(1) error process parameters for n=300 (straight line + dots), n=500 (dotted line), and n=1000 (straight line). A conservative 95% critical region is indicated by two dash dotted lines; b) ASE for model I) and* α*=0.05 versus the AR(1) error process parameters for n=300 (straight line + dots), n=500 (dotted line), and n=1000 (straight line).*



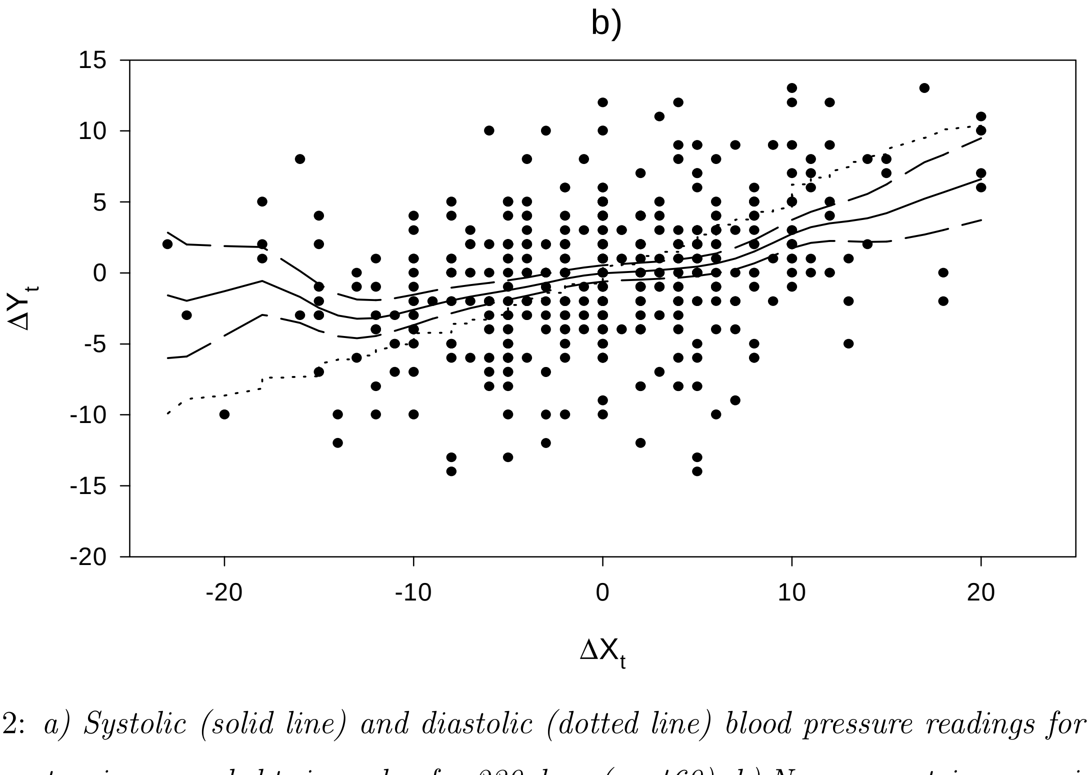


Figure 2: *a) Systolic (solid line) and diastolic (dotted line) blood pressure readings for a mild hypertensive, sampled twice a day for 230 days (n=460); b) Nonparametric regression estimates of the first differences of the blood pressure data; solid line* $= m_n(\Delta X_t)$*; medium dashed lines are 95% confidence bands; dotted line = fitted linear DR model (13).*

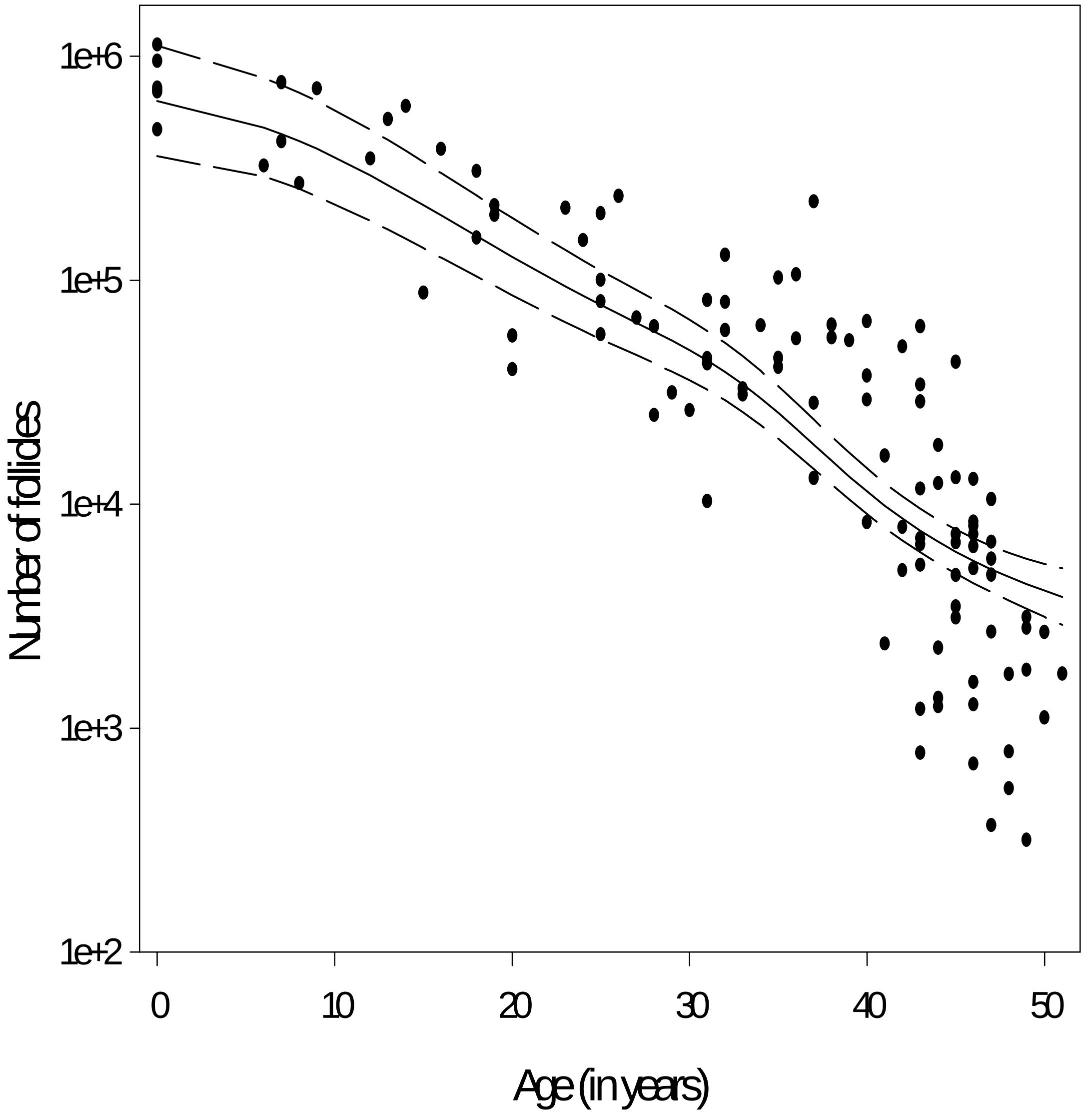


Figure 3: *Ovarian follicle data and nonparametric regression estimates; solid line =* $m_n(X_t)$*; medium dashed lines are 95% confidence bands.*