Zuzana Prášková

Bootstrap in nonstationary autoregression

Kybernetika, Vol. 38 (2002), No. 4, [389]--404

Persistent URL: `http://dml.cz/dmlcz/135473`

KYBERNETIKA — VOLUME 38 (2002), NUMBER 4, PAGES 389-404

BOOTSTRAP IN NONSTATIONARY AUTOREGRESSION[1]

ZUZANA PRÁŠKOVÁ

The first-order autoregression model with heteroskedastic innovations is considered and it is shown that the classical bootstrap procedure based on estimated residuals fails for the least-squares estimator of the autoregression coefficient. A different procedure called wild bootstrap, respectively its modification is considered and its consistency in the strong sense is established under very mild moment conditions.

1. INTRODUCTION

Let $X_1, \dots, X_n$ be observations of a time series satisfying the model

$$X_t = \beta X_{t-1} + Y_t, \quad t = 1, 2, \dots \tag{1}$$

where $|\beta| < 1$ is an unknown parameter, $Y_t, 1 \le t \le n$, are independent random variables with $EY_t = 0$, $\operatorname{Var} Y_t = \sigma_t^2 > 0$ and X_0 is a random variable independent of $Y_1, \dots, Y_n$ such that $EX_0 = 0$, $\operatorname{Var} X_0 = \sigma_0^2 > 0$.

In this paper, we deal with a bootstrap approximation of the distribution of the least-squares estimator of the parameter β. Recently, the problem was solved under the assumption that the innovations Y_t are identically distributed (see e.g. Bose [3], Kreiss and Franke [12], Prášková [16] for $|\beta| < 1$, Basawa et al [1] for $|\beta| > 1$, Datta [6], and Heimann and Kreiss [11] for general β. Ferretti and Romo [9] proposed bootstrap tests for $\beta = 1$ both for independent and autoregressive errors. All the above quoted authors considered a bootstrap procedure based on estimated residuals. Kreiss [13] treated asymptotic properties of this procedure in general stationary autoregression.

However, in case of nonidentically distributed innovations the method need not be consistent (even in a simple linear regression model, see e.g. Liu [14]). We shall show that the bootstrap based on estimated residuals in model (1) generally fails for the least-squares estimator of β. Then we shall consider procedure called wild or external bootstrap and its modification which reflects the heteroskedasticity of data and prove that these procedures consistently estimate the distribution of the

[1]Partially supported by the Grant Agency of the Czech Republic under Grant No. 201/00/0769 and by the Grant MSM 113200008.

least-squares estimator of the parameter β. We shall demonstrate theoretical results in a short simulation study.

2. ASYMPTOTIC RESULTS FOR $\widehat{\beta}$

First, we give some asymptotic results for the least-squares estimator of β in case of nonidentically distributed innovations. Let us introduce the following assumptions:

A1: For some $\delta > 0$ and a positive constant K, $E|Y_t|^{2+\delta} \le K$ for all t, $E|X_0|^{2+\delta} \le K$.

A2: $\frac{1}{n}\sum_{t=1}^{n} \sigma_t^2 \to \sigma^2 > 0$ as $n \to \infty$.

A3: $s_n^2 = \frac{1}{n}\sum_{t=1}^{n} \sigma_t^2 E X_{t-1}^2 \to \overline{\sigma}^2 > 0$ as $n \to \infty$.

Theorem 1. Suppose that assumptions A1 – A3 hold. Let $\widehat{\beta}$ be the least-squares estimator of β based on $X_0, X_1, \ldots, X_n$, i.e.

$$\widehat{\beta} = \frac{\sum_{t=1}^{n} X_t X_{t-1}}{\sum_{t=1}^{n} X_{t-1}^2}. \tag{2}$$

Then

(i) $\widehat{\beta}$ is strongly consistent, i.e. $\widehat{\beta} \to \beta$ a.s. as $n \to \infty$;

(ii) the asymptotic distribution of $\sqrt{n}(\widehat{\beta} - \beta)$ is $\mathcal{N}(0, \Delta^2)$, where

$$\Delta^2 = \frac{(1-\beta^2)^2 \overline{\sigma}^2}{\sigma^4}. \tag{3}$$

Proof. With $Y_0 := X_0$ we can write $X_t = \sum_{j=0}^{t} \beta^j Y_{t-j}$ and utilizing the Minkowski inequality we get

$$\left(E|X_t|^{2+\delta}\right)^{\frac{1}{2+\delta}} \le \sum_{j=0}^{t} \left(|\beta|^{(2+\delta)j} E|Y_{t-j}|^{2+\delta}\right)^{\frac{1}{2+\delta}} \le \sum_{j=0}^{t} |\beta|^j K^{\frac{1}{2+\delta}}$$

which means that $E|X_t|^{2+\delta} \le M$ for a positive constant M and $t \ge 0$. Notice that

$$\widehat{\beta} - \beta = \frac{\sum_{t=1}^{n} X_{t-1} Y_t}{\sum_{t=1}^{n} X_{t-1}^2}. \tag{4}$$

Further, $E(X_{t-1}Y_t | \mathcal{F}_{t-1}) = 0$, where $\mathcal{F}_t = \sigma\{Y_0, Y_1, \ldots, Y_t\}$ for $t \ge 0$ is the σ-algebra generated by $Y_0, Y_1, \ldots, Y_t$. Thus, $\{X_{t-1}Y_t\}$ is a martingale differences sequence. Next, Var $(X_{t-1}Y_t) = E X_{t-1}^2 Y_t^2 \le C$, where C is a constant, and according to the strong law of large numbers for martingale difference sequences (see Davidson [7], Theorem 20.11)

$$\frac{1}{n}\sum_{t=1}^{n} X_{t-1}Y_t \to 0 \quad \text{as } n \to \infty \text{ a.s.} \tag{5}$$

In the following, we prove that

$$\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 \to \frac{\sigma^2}{1-\beta^2} \quad \text{as } n \to \infty \text{ a.s.} \tag{6}$$

From (1) we get $X_t^2 = Y_t^2 + 2\beta X_{t-1}Y_t + \beta^2 X_{t-1}^2$ and

$$(1-\beta^2)\frac{1}{n}\sum_{t=1}^{n} X_{t-1}^2 = \frac{1}{n}(X_0^2 - X_n^2) + \frac{1}{n}\sum_{t=1}^{n} Y_t^2 + 2\beta\frac{1}{n}\sum_{t=1}^{n} X_{t-1}Y_t. \tag{7}$$

From A1 and the strong law of large numbers we have

$$\frac{1}{n}\sum_{t=1}^{n} Y_t^2 - \frac{1}{n}\sum_{t=1}^{n} \sigma_t^2 \to 0 \quad \text{a.s.} \tag{8}$$

Combining this with (5), A1 and A2, we get (6) and assertion (i). Since

$$\frac{1}{ns_n}\sum_{t=1}^{n} X_{t-1}^2 \to \frac{\sigma^2}{\bar{\sigma}(1-\beta^2)} \quad \text{a.s.}$$

and thus in probability, we prove (ii) when we show that $\sum_{t=1}^{n} X_{t-1}Y_t/(s_n\sqrt{n})$ has asymptotically $\mathcal{N}(0,1)$ distribution. It suffices to check that the following conditions for martingale central limit theorem are satisfied (see Brown [5]):

$$\frac{\sum_{t=1}^{n} E\left((X_{t-1}Y_t)^2|\mathcal{F}_{t-1}\right)}{\sum_{t=1}^{n} E(X_{t-1}Y_t)^2} = \frac{\sum_{t=1}^{n} X_{t-1}^2\sigma_t^2}{ns_n^2} \to 1 \quad \text{as } n \to \infty \text{ in probability,} \tag{9}$$

$$\frac{1}{ns_n^2}\sum_{t=1}^{n} E\left((X_{t-1}Y_t)^2 I\{|X_{t-1}Y_t| > \epsilon\sqrt{n}s_n\}\right) \to 0 \quad \text{as } n \to \infty \tag{10}$$

for all $\epsilon > 0$.

We prove that

$$\frac{1}{n}\sum_{t=1}^{n} \sigma_t^2 X_{t-1}^2 - \frac{1}{n}\sum_{t=1}^{n} \sigma_t^2 E X_{t-1}^2 \to 0 \quad \text{a.s.} \tag{11}$$

Then, because of A3, condition (9) will be satisfied.

Denote $\xi_t = \sigma_t^2(X_{t-1}^2 - EX_{t-1}^2)$. We shall show that for $1 < p < 2$, $\{\xi_t\}$ is the L_p-mixingale of size -1, where L_p denotes the usual norm space of random variables with finite moments of order p (see Davidson [7], Chapter 16.1 for the definition of mixingales).

Obviously, $E(\xi_t|\mathcal{F}_{t+s}) = \xi_t$ a. s. for any $s \geq 0$, thus $||\xi_t - E(\xi_t|\mathcal{F}_{t+s})||_p = 0$. Further, we have $X_t = \sum_{j=0}^{k-1} \beta^j Y_{t-j} + \beta^s X_{t-s}$ and $E(\xi_t|\mathcal{F}_{t-s}) = \sigma_t^2 |\beta|^{2s}(X_{t-s}^2 - EX_{t-s}^2)$, thus

$$||E(\xi_t|\mathcal{F}_{t-s})||_p = \sigma_t^2 |\beta|^{2s} ||X_{t-s}^2 - EX_{t-s}^2||_p \leq c|\beta|^{2s},$$

where c is a positive constant independent of t. Then, from the strong law of large numbers for mixingales, (11) holds a. s. according to Theorem 20.16 in Davidson [7].

Finally, we have

$$\frac{1}{ns_n^2} \sum_{t=1}^{n} E\left((X_{t-1}Y_t)^2 I\{|X_{t-1}Y_t| > \epsilon\sqrt{n}s_n\}\right)$$

$$\leq \frac{1}{\epsilon^\delta n^{1+\delta} s_n^{2+\delta}} \sum_{t=1}^{n} E|X_{t-1}|^{2+\delta} E|Y_t|^{2+\delta} \leq M \frac{1}{\epsilon^\delta n^{1+\delta} s_n^{2+\delta}} \sum_{t=1}^{n} E|Y_{t-1}|^{2+\delta} \quad (12)$$

from which(10) easily follows. □

Remark 1. In case that $\sigma_t^2 \equiv \sigma^2$ (asymptotic weak stationarity), Assumption A2 holds trivially and A3 holds with $\bar{\sigma}^2 = \sigma^4(1-\beta^2)^{-1}$. Thus, the results of Theorem 1 coincide with those for stationary $AR(1)$ process (see Brockwell and Davis [4], Chapters 7, 8.) Some other generalizations of the assumption of i.i.d. innovations in autoregressive models of a finite order $p \geq 1$ were considered and central limit theorems were established (see e. g. Hall and Heyde [10], Dürr and Loges [8], Tjøstheim and Paulsen [18] or Basu and Roy [2] among others.)

Notice that the asymptotic variance Δ^2 depends on the parameter β (usually unknown) and on limiting values σ^2 and $\overline{\sigma}^2$ which are also unknown. In next sections we shall deal with the bootstrap approximation of $\sqrt{n}(\widehat{\beta} - \beta)$.

3. BOOTSTRAP BASED ON ESTIMATED RESIDUALS

Let $X_0, \ldots, X_n$ be observations and $\widehat{\beta}$ be the least-squares estimator of the parameter β. Put

$$r_t = X_t - \widehat{\beta} X_{t-1}, t = 1, \ldots, n, \quad \bar{r} = \frac{1}{n} \sum_{t=1}^{n} r_t \quad (13)$$

and consider centered estimated residuals $\widehat{Y}_t = r_t - \bar{r}, t = 1, \ldots, n$. Let F_n be the empirical distribution function based on $\widehat{Y}_1, \ldots, \widehat{Y}_n$ and $Y_0^*, Y_1^*, \ldots, Y_n^*$ be i.i.d. with the distribution function F_n.

Define $X_0^* = Y_0^*$ and generate bootstrap values

$$X_t^* = \widehat{\beta} X_{t-1}^* + Y_t^*, \quad t = 1, \ldots, n. \quad (14)$$

Let

$$\widehat{\beta}^* = \frac{\sum_{t=1}^n X_{t-1}^* X_t^*}{\sum_{t=1}^n X_{t-1}^{*2}}$$

be the bootstrap counterparts of $\widehat{\beta}$.

In the case of i.i.d. innovations Y_t the above bootstrap procedure is consistent, i. e. the bootstrap distribution of $\sqrt{n}(\widehat{\beta}^* - \widehat{\beta})$ converges to the true distribution of $\sqrt{n}(\widehat{\beta} - \beta)$ (see e. g. Bose [3], Kreiss and Franke [12], Prášková [16], Kreiss [13].)

However, when Y_t are independent with zero mean but different variances, the method becomes inconsistent. We shall show it in the next theorem.

Theorem 2. Under assumptions A1 – A3, as $n \to \infty$,

$$\sup_x |P^*(\sqrt{n}(\widehat{\beta}^* - \widehat{\beta}) < x) - \Phi(x/\sqrt{1-\beta^2})| \to 0 \quad \text{a. s.}$$

where P^* denotes the bootstrap probability and Φ is the distribution function of $\mathcal{N}(0,1)$.

Proof. Let E^*, Var^* be the expectation, respectively the variance related to P^*. Then

$$\begin{aligned} E^* Y_1^* &= \frac{1}{n}\sum_{t=1}^n \widehat{Y}_t = \frac{1}{n}\sum_{t=1}^n (r_t - \bar{r}) = 0 \\ \operatorname{Var}^* Y_1^* &= \frac{1}{n}\sum_{t=1}^n \widehat{Y}_t^2 = \frac{1}{n}\sum_{t=1}^n r_t^2 - \bar{r}^2 := \sigma^{*2}. \end{aligned} \tag{15}$$

Since $r_t = X_t - \widehat{\beta} X_{t-1} = Y_t - (\widehat{\beta} - \beta) X_{t-1}$, we can deduce from A1, the strong law of large numbers for $\{Y_t\}$ and from the consistency of $\widehat{\beta}$ that $\bar{r} \to 0$ a. s. Similarly, from (5), (6), (8) and A2 we get that $\frac{1}{n}\sum_{t=1}^n r_t^2$ is asymptotically σ^2 a. s. Thus, we can conclude that

$$\sigma^{*2} \to \sigma^2 \quad \text{as } n \to \infty \quad \text{a. s.} \tag{16}$$

From the relation $X_t^* = \sum_{j=0}^t \widehat{\beta}^j Y_{t-j}^*$ and the independence of $Y_0^*, \dots, Y_n^*$ it follows

$$\frac{1}{n}\sum_{t=1}^n E^* X_{t-1}^{*2} = \frac{\sigma^{*2}}{1-\widehat{\beta}^2}\left[1 - \frac{1}{n}\cdot\frac{1-\widehat{\beta}^{2n}}{1-\widehat{\beta}^2}\right]$$

and from (16) and the strong consistency of $\widehat{\beta}$ we get

$$\frac{1}{n}\sum_{t=1}^n E^* X_{t-1}^{*2} \to \frac{\sigma^2}{1-\beta^2} \quad \text{a. s.} \tag{17}$$

Similarly,

$$s_n^{*2} = \frac{1}{n}\sum_{t=1}^{n} E^* X_{t-1}^{*2} E^* Y_t^{*2} = \sigma^{*2}\frac{1}{n}\sum_{t=1}^{n} E^* X_{t-1}^{*2} \to \frac{\sigma^4}{1-\beta^2} \quad \text{a.s.} \tag{18}$$

When we apply a version of Marcinkiewicz strong law of large numbers (Lemma 1 in Liu [14]) we get with δ from Assumption 1

$$\frac{1}{n}\sum_{t=1}^{n} |Y_t|^{2+\frac{\delta}{2}} - \frac{1}{n}\sum_{t=1}^{n} E|Y_t|^{2+\frac{\delta}{2}} \to 0 \quad \text{a.s.}$$

hence,

$$\frac{1}{n}\sum_{t=1}^{n} |Y_t|^{2+\frac{\delta}{2}} = O(1) \quad \text{a.s.} \tag{19}$$

From the same Lemma we also get that

$$\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^{n} |Y_t|^{2+\delta} \to 0 \quad \text{a.s.} \tag{20}$$

Hence, $\max_{1\le t\le n} |Y_t| = o(n^{\frac{1}{2}})$ a.s. and thus $\max_{1\le t\le n} |X_t| = o(n^{\frac{1}{2}})$. From here and from (6) we have $\frac{1}{n}\sum_{t=1}^{n} |X_{t-1}|^{2+\frac{\delta}{2}} = o(n^{\frac{\delta}{4}})$ a.s. When we use Theorem 20.11 in Davidson [7] with $p = 2$ and $a_n = n^{\frac{1}{2}+\varepsilon}$ for some $\varepsilon > 0$, we can see that

$$\widehat{\beta} - \beta = o(n^{-\frac{1}{2}+\varepsilon}) \quad \text{a.s.} \tag{21}$$

and with properly chosen ε

$$\frac{1}{n}\sum_{t=1}^{n} |r_t|^{2+\frac{\delta}{2}} \le 4\left(\frac{1}{n}\sum_{t=1}^{n} |Y_t|^{2+\frac{\delta}{2}} + |\widehat{\beta}-\beta|^{2+\frac{\delta}{2}}\frac{1}{n}\sum_{t=1}^{n} |X_{t-1}|^{2+\frac{\delta}{2}}\right) = O(1) \quad \text{a.s.} \tag{22}$$

It means that with δ as in Assumption 1,

$$E^*|Y_1^*|^{2+\frac{\delta}{2}} = \frac{1}{n}\sum_{t=1}^{n} |r_t - \bar{r}|^{2+\frac{\delta}{2}} = O(1) \quad \text{a.s.} \tag{23}$$

In a similar way we obtain

$$\frac{1}{n^2}\sum_{t=1}^{n} |Y_t|^4 \to 0 \quad \text{a.s.} \tag{24}$$

and thus

$$E^*|Y_1^*|^4 = \frac{1}{n}\sum_{t=1}^{n} |r_t - \bar{r}|^4 = o(n) \quad \text{a.s.} \tag{25}$$

Now, let us write

$$\frac{\sqrt{n}(\widehat{\beta}^* - \widehat{\beta})}{\sqrt{1-\widehat{\beta}^2}} = \left(\frac{1}{s_n^*\sqrt{n}}\sum_{t=1}^n X_{t-1}^* Y_t^*\right)\left(\frac{\sqrt{1-\widehat{\beta}^2}}{n s_n^*}\sum_{t=1}^n X_{t-1}^{*2}\right)^{-1}.$$

According to an extension of Lemma 1 in Michel and Pfanzagl [15] (see Basu and Roy [2], Lemma 2.1), for any $\epsilon > 0$ and real V there exists $0 < c < 1$ such that

$$\sup_x \left| P^*\left(\sqrt{n}\frac{\widehat{\beta}^* - \widehat{\beta}}{\sqrt{1-\widehat{\beta}^2}} \le x\right) - \Phi(x)\right| \le \sup_x \left| P^*\left(\frac{1}{s_n^*\sqrt{n}}\sum_{t=1}^n X_{t-1}^* Y_t^* < x\right) - \Phi(x)\right|$$

$$+P^*\left(\left|\frac{\sqrt{1-\widehat{\beta}^2}}{n s_n^*}\sum_{t=1}^n X_{t-1}^{*2} - V\right| > \epsilon\right) + \epsilon + c|V-1|. \tag{26}$$

Put

$$V = \frac{\sqrt{1-\widehat{\beta}^2}}{n s_n^*}\sum_{t=1}^n E^* X_{t-1}^{*2}.$$

Then, (17), (18) and the strong consistency of $\widehat{\beta}$ yields

$$V - 1 \to 0 \quad \text{a.s.} \tag{27}$$

Further, from the Chebyshev inequality we have

$$P^*\left(\left|\frac{\sqrt{1-\widehat{\beta}^2}}{n s_n^*}\sum_{t=1}^n X_{t-1}^{*2} - V\right| > \epsilon\right) \le \frac{1-\widehat{\beta}^2}{\epsilon^2 s_n^{*2}} E^*\left[\frac{1}{n}\sum_{t=1}^n (X_{t-1}^{*2} - E^* X_{t-1}^{*2})\right]^2$$

and since

$$(1-\widehat{\beta}^2)\frac{1}{n}\sum_{t=1}^n (X_{t-1}^{*2} - E^* X_{t-1}^{*2})$$

$$= \left[\frac{1}{n}((X_0^{*2} - E^* X_0^{*2}) - (X_n^{*2} - E^* X_n^{*2})) + \frac{1}{n}\sum_{t=1}^n (Y_t^{*2} - \sigma^{*2}) + 2\widehat{\beta}\frac{1}{n}\sum_{t=1}^n X_{t-1}^* Y_t^*\right]$$

we can easily check that

$$E^*\left[\frac{1}{n}\sum_{t=1}^n (X_{t-1}^{*2} - E^* X_{t-1}^{*2})\right]^2 = \frac{1}{(1-\widehat{\beta}^2)^2}\left[\frac{1}{n}(E^* Y_1^{*4} - \sigma^{*4})\right] + \frac{4}{n}\widehat{\beta}^2 s_n^{*2} + Z_n$$

where Z_n is $o(1)$ a.s.

From here and from (16), (18) and (25) we can conclude that

$$E^*\left[\frac{1}{n}\sum_{t=1}^n (X_{t-1}^{*2} - E^* X_{t-1}^{*2})\right]^2 \to 0 \quad \text{a.s.} \tag{28}$$

and thus the second term on the right-hand side of (26) tends to zero a. s.

Notice that with (28) the bootstrap version of (11) and (9) is satisfied in P^*-probability.

Finally,

$$\frac{1}{ns_n^{*2}} \sum_{t=1}^{n} E^*((X_{t-1}^* Y_t^*)^2 I\{|X_{t-1}^* Y_t^*| > \epsilon s_n^* \sqrt{n}\})$$
$$\leq \frac{1}{\epsilon^{\frac{\delta}{2}} n^{1+\frac{\delta}{4}} (s_n^*)^{2+\frac{\delta}{2}}} E^*|Y_1^*|^{2+\frac{\delta}{2}} \sum_{t=1}^{n} E^*|X_{t-1}^*|^{2+\frac{\delta}{2}} \tag{29}$$

and since

$$E^*|X_{t-1}^*|^{2+\frac{\delta}{2}} \leq E^*|Y_1^*|^{2+\frac{\delta}{2}} \left(\frac{1-|\widehat{\beta}|^t}{1-|\widehat{\beta}|} \right)^{2+\frac{\delta}{2}}$$

which follows from the Minkowski inequality, we can conclude, using (23) and (18) that the right-hand side of (29) is asymptotically zero a. s. and the bootstrap version of (10) is satisfied. It means that

$$\sup_x \left| P^* \left(\frac{1}{s_n^* \sqrt{n}} \sum_{t=1}^{n} X_{t-1}^* Y_t^* < x \right) - \Phi(x) \right| \to 0 \quad \text{a. s.}$$

which together with the strong consistency of $\widehat{\beta}$ concludes the proof. □

We can see that the asymptotic variance of $\sqrt{n}(\widehat{\beta}^* - \widehat{\beta})$ differs from that of $\sqrt{n}(\widehat{\beta} - \beta)$ given in (3). The bootstrap scheme (14) does not reflect the heteroskedasticity of the original data because it works with the innovations Y_t^* which are (conditionally on $X_0, \dots, X_n$) independent and identically distributed.

Another bootstrap procedure can solve the problem of heteroskedasticity.

4. WILD BOOTSTRAP

In Kreiss [13] the bootstrap procedure is discussed, which mimics a procedure called wild bootstrap proposed for regression models with heteroskedastic errors (see e. g. Wu [19] or Liu [14]).

With residuals $r_t = X_t - \widehat{\beta} X_{t-1}$, where $\widehat{\beta}$ is given in (2), the bootstrap innovations are generated as

$$Y_t^w = r_t K_t, \quad t = 1, \dots, n \tag{30}$$

where K_t are i.i.d. random variables with zero mean and the unit variance, independent of $X_0, \dots, X_n$. Given observations $X_0, \dots, X_n$, the bootstrap observations are generated to satisfy

$$X_t^w = \widehat{\beta} X_{t-1} + Y_t^w, \quad t = 1, \dots, n \tag{31}$$

and the corresponding bootstrap estimator of β is then defined as the least-squares estimator in the regression model

$$X_t^w = \beta X_{t-1} + Y_t^w, \quad t = 1\dots, n$$

with constant regressors X_{t-1}, i. e.

$$\widehat{\beta}^w = \frac{\sum_{t=1}^n X_{t-1} X_t^w}{\sum_{t=1}^n X_{t-1}^2}. \tag{32}$$

In Kreiss [13], the procedure is considered and its consistency (in probability) is studied in models with i.i.d. errors, respectively in stationary models with conditional heteroskedasticities. Here we give a proof of the strong consistency of the procedure in nonstationary model (1) under weaker moment conditions than in Kreiss [13]. Instead of Assumption A1 let us assume

A1': For some $\delta > 0$ and a positive constant M, $E|Y_t|^{3+\delta} \le M$ for all t, $E|X_0|^{3+\delta} \le M$. Random variables $K_1, \dots, K_n$ are i.i.d. with zero mean, unit variance and finite moment of order $2+\delta', \delta' \ge \delta$.

Theorem 3. Under assumptions A1', A2, A3, as $n \to \infty$,

$$\sup_x |P^w(\sqrt{n}(\widehat{\beta}^w - \widehat{\beta}) < x) - \Phi(x/\Delta))| \to 0 \quad \text{a. s.}$$

where P^w is the conditional probability given $X_0, \dots, X_n$ and Δ is given in (4).

Proof. Notice that

$$\sqrt{n}(\widehat{\beta}^w - \widehat{\beta}) = \frac{\frac{1}{\sqrt{n}} \sum_{t=1}^n X_{t-1} Y_t^w}{\frac{1}{n} \sum_{t=1}^n X_{t-1}^2}$$

is a linear combination of independent random variables Y_t^w for which

$$\begin{aligned} E^w Y_t^w &= E(r_t K_t | X_0, \dots, X_n) = 0, \\ \mathrm{Var}^w Y_t^w &= r_t^2 \operatorname{Var} K_t = r_t^2, \\ E^w |Y_t^w|^{2+\delta} &= |r_t|^{2+\delta} E|K_t|^{2+\delta} = c|r_t|^{2+\delta} \end{aligned} \quad . \tag{33}$$

where $c = E|K_1|^{2+\delta} < \infty$.

Since (6) remains valid under assumptions of Theorem 3, it suffices to prove the asymptotic normality of $\sum_{t=1}^n Z_{tn}$, where $Z_{tn} = X_{t-1} Y_t^w / \sqrt{n}$. Let us denote

$$B_n^2 = \sum_{t=1}^n \mathrm{Var}^w Z_{tn} = \frac{1}{n} \sum_{t=1}^n X_{t-1}^2 r_t^2. \tag{34}$$

Then we have

$$B_n^2 = \frac{1}{n} \sum_{t=1}^n X_{t-1}^2 Y_t^2 - 2(\widehat{\beta} - \beta) \frac{1}{n} \sum_{t=1}^n X_{t-1}^3 Y_t + (\widehat{\beta} - \beta)^2 \frac{1}{n} \sum_{t=1}^n X_{t-1}^4. \tag{35}$$

Since $X_{t-1}^2(Y_t^2 - EY_t^2)$ and $X_{t-1}^3 Y_t$ are martingale differences, we can easily check, using Assumptions A1', A3 and (11), that the a. s. limit of the first term on the right-hand side of (35) is $\overline{\sigma}^2$ and the second term tends to 0 a. s. Under Assumption A1', $\max_{1\le t\le n}|Y_t| = o(n^{\frac{1}{3}})$ a. s. and the same holds for $\max_{1\le t\le n}|X_t|$. Combining this with (21) we obtain that the last term on the right-hand side of (35) is $o(n^{-\frac{1}{3}+\epsilon})$ for some $\epsilon > 0$, thus, $B_n^2 \to \overline{\sigma}^2$ a. s.

To verify the Feller–Lindeberg condition, write

$$\frac{1}{B_n^2}\sum_{t=1}^n E^w\left[\left(\frac{X_{t-1}Y_t^w}{\sqrt{n}}\right)^2 I\left\{\left|\frac{X_{t-1}Y_t^w}{\sqrt{n}}\right| > \epsilon B_n\right\}\right]$$
$$\le c\frac{1}{\epsilon^\delta B_n^{2+\delta}}\cdot\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n E^w|X_{t-1}Y_t^w|^{2+\delta} \le \frac{1}{\epsilon^\delta B_n^{2+\delta}}\cdot\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n |X_{t-1}r_t|^{2+\delta} \quad (36)$$

further,

$$\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n |X_{t-1}r_t|^{2+\delta}$$
$$\le 4\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n |X_{t-1}|^{2+\delta}|Y_t|^{2+\delta} + 4(\widehat{\beta}-\beta)^{2+\delta}\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n |X_{t-1}|^{4+2\delta}. \quad (37)$$

The second term on the right-hand side of (37) tends to zero a. s. similarly as the last term in (35) while for the first one we get

$$\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n |X_{t-1}|^{2+\delta}(|Y_t|^{2+\delta} - E|Y_t|^{2+\delta}) \to 0 \quad \text{a. s.}$$

according to the strong law of large numbers for martingale differences and

$$\frac{1}{n^{1+\frac{\delta}{2}}}\sum_{t=1}^n |X_{t-1}|^{2+\delta}E|Y_t|^{2+\delta} \to 0 \quad \text{a. s.}$$

which follows from (6) and (20). The proof is finished. □

Corollary 1. Under assumptions of Theorem 3, as $n \to \infty$,

$$\sup_x |P^w(\sqrt{n}(\widehat{\beta}^w - \widehat{\beta}) < x) - P(\sqrt{n}(\widehat{\beta} - \beta) < x)| \to 0 \quad \text{a. s.}$$

5. MODIFIED WILD BOOTSTRAP

Bootstrap procedure (31) can be modified in the following way. Consider again the bootstrap innovations Y_t^w defined by (30) and generate bootstrap observations as follows. Put $X_0^{*w} = 0$ and further generate

$$X_t^{*w} = \widehat{\beta}X_{t-1}^{*w} + Y_t^w, \; t = 1, \dots, n, \quad (38)$$

$\widehat{\beta}$ is defined by (2). Notice that procedure (38) generates bootstrap observations that follow the same model as original observations.

Let $\widehat{\beta}^{*w}$ be the least-squares estimator of autoregressive parameter in the bootstrap model (38), i. e.

$$\widehat{\beta}^{*w} = \frac{\sum_{t=1}^{n} X_{t-1}^{*w} Y_t^w}{\sum_{t=1}^{n} (X_{t-1}^{*w})^2}.$$

To prove the strong consistency of $\widehat{\beta}^{*w}$ we need to replace Assumption A1' by

A1": For some $\delta > 0$ and a positive constant M, $E|Y_t|^{4+\delta} \le M$ for all t, $E|X_0|^{4+\delta} \le M$. Random variables $K_1, \dots, K_n$ are i.i.d. with zero mean, unit variance and finite moment of order 4.

Remark 2. Consistency of $\widehat{\beta}^{*w}$ was studied by Kreiss [13] in stationary autoregression and in Prášková [17] for nonstationary model (1) under strong moment conditions.

Theorem 4. Under assumptions A1", A2, A3, as $n \to \infty$,

$$\sup_x |P^w(\sqrt{n}(\widehat{\beta}^{*w} - \widehat{\beta}) < x) - \Phi(x/\Delta))| \to 0 \quad \text{a. s.}$$

where P^w is the conditional probability given $X_0, \dots, X_n$ and Δ is given in (4).

Proof. We will proceed similarly as in the proof of Theorem 2. From (38) we have

$$\begin{aligned} X_0^{*w} &= 0, \\ X_{t-1}^{*w} &= \sum_{k=0}^{t-2} \widehat{\beta}^k Y_{t-1-k}^w = \sum_{j=1}^{t-1} \widehat{\beta}^{t-1-j} Y_j^w \quad \text{for } t \ge 2, \end{aligned}$$

hence,

$$\begin{aligned} \frac{1}{n}\sum_{t=2}^{n} E^w(X_{t-1}^{*w})^2 &= \frac{1}{n}\sum_{t=2}^{n}\sum_{j=1}^{t-1} \widehat{\beta}^{2(t-1-j)} r_j^2 = \frac{1}{n}\sum_{t=1}^{n-1} r_t^2 \sum_{k=t+1}^{n} \widehat{\beta}^{2(k-t-1)} \\ &= \frac{1}{1-\widehat{\beta}^2}\left[\frac{1}{n}\sum_{t=1}^{n-1} r_t^2 - \frac{1}{n}\sum_{t=1}^{n-1} r_t^2 \widehat{\beta}^{2(n-t)}\right]. \end{aligned} \tag{39}$$

Now, we can see that the first term in the last equality in (39) is asymptotically $\frac{\sigma^2}{1-\beta^2}$ a. s. and

$$\frac{1}{n}\sum_{t=1}^{n-1} r_t^2 \widehat{\beta}^{2(n-t)} \le \max_{1 \le t \le n-1} |r_t|^2 \frac{1}{n} \cdot \frac{\widehat{\beta}^2 - \widehat{\beta}^{2n}}{1-\widehat{\beta}^2} = o(1) \quad \text{a. s.}$$

since $\max|r_t| = o(n^{\frac{1}{4}})$ under A1". Thus,

$$\frac{1}{n}\sum_{t=2}^{n} E^w(X_{t-1}^{*w})^2 \to \frac{\sigma^2}{1-\beta^2} \quad \text{a.s.} \tag{40}$$

Similarly,

$$s_n^{w2} = \frac{1}{n}\sum_{t=1}^{n} E^w(X_{t-1}^{*w}Y_t^w)^2 = \frac{1}{n}\sum_{t=2}^{n} r_t^2 \sum_{j=1}^{t} \widehat{\beta}^{2(t-1-j)} r_j^2. \tag{41}$$

When we insert $r_t = Y_t - (\widehat{\beta} - \beta)X_{t-1}$ into (41) and split the factors, then, using Assumptions A1", A3, the strong law of large numbers for martingale differences $(Y_t^2 - EY_t^2)\sum_{j=1}^{t-1} \beta^{2(t-j-1)} Y_j^2$, similar considerations that led to (11) and the strong consistency of $\widehat{\beta}$, we get that

$$\frac{1}{n}\sum_{t=2}^{n} Y_t^2 \sum_{j=1}^{t} \widehat{\beta}^{2(t-j-1)} Y_j^2 \to \overline{\sigma}^2 \quad \text{a.s.}$$

and after very careful analysis of the other terms that appear on the right-hand side of (41) we obtain

$$s_n^{w2} \to \overline{\sigma}^2 \quad \text{a.s.} \tag{42}$$

Now, analogously to (26), we get

$$\sup_x \left| P^w\left(\frac{\sqrt{n}(\widehat{\beta}^{*w} - \widehat{\beta})}{\Delta} < x \right) - \Phi(x) \right| \le \sup_x \left| P^w\left(\frac{1}{s_n^w\sqrt{n}} \sum_{t=1}^{n} X_{t-1}^{*w} Y_t^w < x \right) - \Phi(x) \right|$$
$$+ P^w\left(\left| \frac{\Delta}{s_n^w}\frac{1}{n}\sum_{t=1}^{n}(X_{t-1}^{*w})^2 - V \right| > \epsilon \right) + \epsilon + c|V-1| \tag{43}$$

where

$$V = \frac{\Delta}{s_n^w}\frac{1}{n}\sum_{t=1}^{n} E^w(X_{t-1}^{*w})^2 \to 1 \quad \text{a.s.}$$

which follows from (40) and (41). Further, repeating considerations that led to (28) with the only modification that $\mathrm{Var}^w(Y_t^w)^2 = r_t^4(EK_1^4 - 1)$ we get under A1" by using the Chebyshev inequality that

$$P^w\left(\left| \frac{\Delta}{s_n^w}\frac{1}{n}\sum_{t=1}^{n}(X_{t-1}^{*w})^2 - V \right| > \epsilon \right) \to 0 \quad \text{a.s.} \tag{44}$$

It remains to prove that a bootstrap analogy of (9) and (10) holds true. Due to (41) and analogously to (11) it suffices to prove that

$$P^w\left(\left| \frac{1}{n}\sum_{t=2}^{n} r_t^2 (X_{t-1}^{*w})^2 - \frac{1}{n}\sum_{t=2}^{n} r_t^2 E^w(X_{t-1}^{*w})^2 \right| > \epsilon \right) \to 0 \quad \text{a.s.} \tag{45}$$

and that

$$\frac{1}{s_n^{w2}}\frac{1}{n}\sum_{t=1}^{n} E^w[(X_{t-1}^{*w}Y_t^w)^2 I\{|X_{t-1}^{*w}Y_t^w| > \epsilon\sqrt{n}s_n^w\}] \to 0 \quad \text{a.s.} \tag{46}$$

To prove (45), observe that

$$\frac{1}{n}\sum_{t=2}^{n} r_t^2((X_{t-1}^{*w})^2 - E^w(X_{t-1}^{*w})^2) = S_1 + S_2$$

where

$$\begin{aligned} S_1 &= \frac{1}{n}\sum_{t=1}^{n-1}(Y_t^{w2} - r_t^2)c_t, \\ S_2 &= \frac{2}{n}\sum_{t=3}^{n} r_t^2 \sum_{j=0}^{t-3}\widehat{\beta}^{2j}V_{t-j-1}^w = \frac{2}{n}\sum_{t=2}^{n-1} c_t V_t^w, \\ c_t &= \sum_{j=t+1}^{n} r_j^2 \widehat{\beta}^{2(j-t-1)}, \\ V_t^w &= Y_t^w \sum_{k=1}^{t-1}\widehat{\beta}^k Y_{t-k}^w. \end{aligned}$$

Hence, from the Chebyshev inequality

$$P^w\left(\left|\frac{1}{n}\sum_{t=2}^{n} r_t^2((X_{t-1}^{*w})^2 - E^w(X_{t-1}^{*w})^2)\right| > \epsilon\right) \le \frac{1}{\epsilon^2}(2E^w S_1^2 + 2E^w S_2^2).$$

Next,

$$E^w S_1^2 = \frac{1}{n^2}\sum_{t=1}^{n-1} c_t^2 E^w(Y_t^{w2} - r_t^2)^2 = (EK_1^4 - 1)\frac{1}{n^2}\sum_{t=1}^{n-1} c_t^2 r_t^4.$$

Obviously, under A1", $\max_{1\le t\le n-1}|c_t| = o(n^{\frac{1}{2}})$ a. s. and $\frac{1}{n}\sum_{t=1}^{n} r_t^4$ remains bounded, thus, $E^w S_1^2 \to 0$ a. s.

Similarly, we find that

$$E^w S_2^2 = \frac{4}{n^2}\sum_{t=2}^{n-1} c_t^2 r_t^2 \sum_{k=1}^{t-1}\widehat{\beta}^{2k} r_{t-k}^2 = \frac{o(n)}{n}\cdot\frac{1}{n}\sum_{t=2}^{n-1} r_t^2 \sum_{k=1}^{t-1}\widehat{\beta}^{2k} r_{t-k}^2 \to 0 \quad \text{a.s.}$$

and thus (45) holds.

For the proof of (46) let us write

$$\begin{aligned} &\frac{1}{s_n^{w2}}\frac{1}{n}\sum_{t=1}^{n} E^w[(X_{t-1}^{*w}Y_t^w)^2 I\{|X_{t-1}^{*w}Y_t^{*w}| > \epsilon\sqrt{n}s_n^w\}] \\ \le\ & \frac{1}{\epsilon s_n^{w3}\sqrt{n}}\frac{1}{n}\sum_{t=1}^{n} E^w|X_{t-1}^{*w}|^3 E^w|Y_t^w|^3. \end{aligned} \tag{47}$$

Since

$$E^w|Y_t^w|^3 = |r_t|^3 E|K_1|^3 \le c(|Y_t|^3 + |\widehat{\beta} - \beta|^3 |X_{t-1}|^3),$$

$$E^w|X_{t-1}^{*w}|^3 \le c \left(\sum_{j=1}^{t-1} |\widehat{\beta}|^{t-1-j} |Y_j| \right)^3 + c|\widehat{\beta} - \beta|^3 \left(\sum_{j=1}^{t-1} |\widehat{\beta}|^{t-1-j} |X_{j-1}| \right)^3$$

where c is a generic positive constant, we get after some computations, utilizing strong consistency of $\widehat{\beta}$, (21) and the fact that $\max_{1 \le t \le n} |Y_t|$, respectively $\max_{1 \le t \le n} |X_t|$ are of order $o(n^{\frac{1}{4}})$ a. s. under A1", that

$$\frac{1}{n^{\frac{3}{2}}} \sum_{t=1}^{n} E^w |X_{t-1}^{*w}|^3 E^w |Y_t^w|^3 \le c \frac{1}{n^{\frac{3}{2}}} \sum_{t=2}^{n} |Y_t|^3 \left(\sum_{j=1}^{t-1} |\beta|^{t-1-j} |Y_j| \right)^3 + o(1)$$

holds almost surely. Further,

$$\begin{aligned} & \frac{1}{n^{\frac{3}{2}}} \sum_{t=2}^{n} |Y_t|^3 \left(\sum_{j=1}^{t-1} |\beta|^{t-1-j} |Y_j| \right)^3 \\ \le \; & c \frac{\max_{2 \le t \le n} |Y_t|^2}{n^{\frac{1}{2}}} \frac{1}{n} \sum_{t=2}^{n} Y_t^2 \left(\sum_{j=1}^{t-1} |\beta|^{t-1-j} |Y_j| \right)^2 = o(1) \quad \text{a. s.} \end{aligned}$$

which concludes the proof of (46). □

Corollary 2. Under assumptions of Theorem 4, as $n \to \infty$,

$$\sup_x |P^w(\sqrt{n}(\widehat{\beta}^{*w} - \widehat{\beta}) < x) - P(\sqrt{n}(\widehat{\beta} - \beta) < x)| \to 0 \quad \text{a. s.}$$

Remark 3. We presented here results for nonstationary AR(1) process, only. However, we can extend them to a general nonstationary process

$$X_t = \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + Y_t, \quad t \ge 0, \; X_{-1} = \cdots = X_{-p} = 0$$

under the assumption that all the roots of the polynomial $\lambda^p - \beta_1 \lambda^{p-1} - \cdots - \beta_p$ lie inside the unit circle. In such case we can write X_t in the form

$$X_t = \sum_{j=0}^{t} c_j Y_{t-j}, \quad t \ge 0$$

where c_j are coefficients that depend on parameters $\beta_1 \ldots, \beta_p$ and geometrically decay to zero (see e. g. Brockwell and Davis [4], Chapter 3). Then we can obtain consistency results for wild bootstrap procedures by using limit theorems for vector martingale differences.

6. SIMULATIONS

We studied all considered bootstrap procedures numerically. We generated nonstationary process (1) with Y_t being independent and normally distributed, with zero mean and the variances $\sigma_t^2 = 1 + 0.5(-1)^t$ for various values of β and sample sizes n. In Figure, true value of asymptotic mean square error of $\widehat{\beta}$ is drawn (bold line) as well as its estimates by residual based bootstrap (dotted line), wild bootstrap (thin line) and modified wild bootstrap (dashed line) for values of β varying from 0.1 to 0.9 and $n = 100$, $n = 200$. We generated 1000 series for each combination of β and n, and 1000 bootstrap replications in each series. The results show that residual based bootstrap does not work well while wild bootstrap does; it gives somewhat better results than modified wild bootstrap for larger values of β but yields larger standard deviations, especially for small values of β.

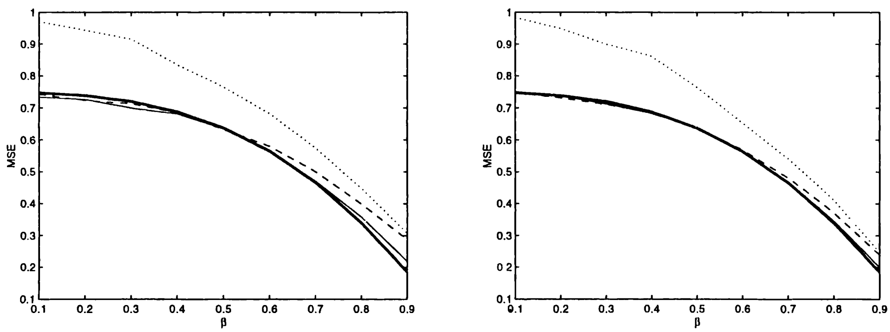

Fig. Estimates of the asymptotic mean square error of $\widehat{\beta}$ by bootstrap for $n = 100$ (left panel) and $n = 200$ observations (right panel).

(Received January 9, 2002.)

REFERENCES

[1] I. V. Basawa, A. K. Mallik, W. P. McCormick, and R. L. Taylor: Bootstrapping explosive autoregressive processes. Ann. Statist. *17* (1989), 1479–1486.

[2] A. K. Basu and S. Sen Roy: On rates of convergence in the central limit theorem for parameter estimation in general autoregressive model. Statistics *21* (1990), 461–470.

[3] A. Bose: Edgeworth correction by bootstrap in autoregression. Ann. Statist. *16* (1988), 1709–1722.

[4] P. J. Brockwell and R. A. Davis: Time Series: Theory and Methods. Springer–Verlag, New York 1987.

[5] B. M. Brown: Martingale central limit theorems. Ann. Math. Statist. *42* (1971), 59–66.

[6] S. Datta: On asymptotic properties of bootstrap for AR(1) processes. J. Statist. Plann. Inference *53* (1996), 361–374.

[7] J. Davidson: Stochastic Limit Theory. Oxford University Press, New York 1994.
[8] D. Dürr and W. Loges: Large deviation results and rates in the central limit theorem for parameter estimators for autoregressive processes. Sankhyā *47* (1985), Series A, 6–24.
[9] N. Ferretti and J. Romo: Unit root bootstrap tests for AR(1) models. Biometrika *83* (1996), 849–860.
[10] P. Hall and C. C. Heyde: Martingale Limit Theory and Its Application. Academic Press, New York 1980.
[11] G. Heimann and J. P. Kreiss: Bootstrapping general first order autoregression. Statist. Probab. Lett. *30* (1996), 87–98.
[12] J. P. Kreiss and J. Franke: Bootstrapping stationary autoregressive moving-average models. J. Time Ser. Anal. *13* (1992), 297–31.
[13] J. P. Kreiss: Asymptotical properties of residual bootstrap for autoregression. Preprint, TU Braunschweig 1997.
[14] R. Y. Liu: Bootstrap procedures under some non-i.i.d. models. Ann. Statist. *16* (1988), 1696–1708.
[15] R. Michel and J. Pfanzagl: The accuracy of the normal approximations for minimum contrast estimate. Z. Wahrsch. Verw. Gebiete *18* (1971), 73–84.
[16] Z. Prášková: A contribution to bootstrapping autoregressive processes. Kybernetika *31* (1995), 359–373.
[17] Z. Prášková: Bootstrapping in autoregression with heteroscedasticities. Working paper, Charles University 1997. Unpublished.
[18] D. Tjøstheim and J. Paulsen: Least squares estimates and order determination procedures for autoregressive process with a time dependent variance. J. Time Ser. Anal. *6* (1985), 117–138.
[19] C. F. J. Wu: Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Statist. *14* (1986), 1261–1295.

Doc. RNDr. Zuzana Prášková, CSc., Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics – Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.
e-mail: praskova@karlin.mff.cuni.cz