KYBERNETIKA — VOLUME 42 (2006), NUMBER 6, PAGES 699 – 709

RATE OF CONVERGENCE FOR A CLASS OF RCA ESTIMATORS

Pavel Vaněček

This work deals with Random Coefficient Autoregressive models where the error process is a martingale difference sequence. A class of estimators of unknown parameter is employed. This class was originally proposed by Schick and it covers both least squares estimator and maximum likelihood estimator for instance. Asymptotic behavior of such estimators is explored, especially the rate of convergence to normal distribution is established.

Keywords: RCA, parameter estimation, rate of convergence

AMS Subject Classification: 60F05, 60G10, 62M10, 91B84

1. INTRODUCTION

In this paper we consider the model of time series called Random Coefficient Autoregressive model of the first order (RCA(1) model). This process is defined as a solution of stochastic equation

$$X_t = (\beta + B_t)\, X_{t-1} + Y_t \ , \tag{1}$$

where β is unknown constant parameter, $\{Y_t, t \in \mathbb{Z}\}$ is an error process with zero expectation and finite constant nonzero variance σ^2, and $\{B_t, t \in \mathbb{Z}\}$ denotes so called random coefficient process with zero expectation and constant nonzero variance σ_B^2 such that $\beta^2 + \sigma_B^2 < 1$.

Both error process and random coefficient process are usually assumed to be mutually independent sequences of independent and identically distributed random variables. In this paper we modify these conditions.

After that, we explore a class of estimators of parameter β proposed by Schick in [10]. We are interested in the rate of convergence of such estimators to normal distribution.

Let us specify the RCA(1) model by making a few assumptions:

A1: $\{B_t, t \in \mathbb{Z}\}$ is a sequence of independent and identically distributed random variables, $\{Y_t, t \in \mathbb{Z}\}$ is an ergodic and strictly stationary sequence of random variables, $\{B_t\}$ and $\{Y_t\}$ are mutually independent.

A2: $\{Y_t, t \in \mathbb{Z}\}$ is a martingale difference sequence with respect to σ-field $\mathcal{F}_t = \sigma(B_s, Y_s; s \le t)$, that is $Y_t \in \mathcal{F}_t$ and $\mathrm{E}[Y_t|\mathcal{F}_{t-1}] = 0$.

It is known that under Assumption A1 there exists strictly stationary and ergodic process $\{X_t, t \in \mathbb{Z}\}$ that satisfies equation (1), see e. g. [8, 12]. This result enables us the following definition.

Definition 1. Process $\{X_t, t \in \mathbb{Z}\}$ will be called RCA(1) process, if $\{X_t\}$ satisfies equation (1) with $\beta^2 + \sigma_B^2 < 1$ and Assumptions A1 and A2.

2. PARAMETER ESTIMATION

We usually employ conditional least squares estimator (LS estimator), its weighted version (WLS estimator), or maximum likelihood estimator (ML estimator) to estimate parameter β in RCA(1) model. Schick in his paper [10] introduced a new class of estimators of parameter β for RCA(1) model. For each measurable function $\phi(x)$, that satisfies $x\phi(x) > 0$ for $x \neq 0$, he defined estimator

$$\widehat{\beta}_n(\phi) = \frac{\sum_{t=1}^{n} \phi(X_{t-1}) \cdot X_t}{\sum_{t=1}^{n} \phi(X_{t-1}) \cdot X_{t-1}} \tag{2}$$

and he proved that, for RCA(1) process with independent and identically distributed error process $\{Y_t\}$, such estimator is strongly consistent and

$$\sqrt{n}(\widehat{\beta}_n(\phi) - \beta) \xrightarrow{D} N(0, V(\phi)) \qquad \text{for } n \to \infty \;,$$

where

$$V(\phi) = \frac{\mathrm{E}\left(\phi^2(X_1) w(X_1)\right)}{\mathrm{E}\left(\phi(X_1) X_1\right)^2} \tag{3}$$

and $w(x) = \sigma^2 + \sigma_B^2 x^2$.

Notice that the choice $\phi(x) = x$ corresponds to LS estimator and $\phi(x) = \frac{x}{w(x)}$ yields WLS estimator which is identical to ML estimator in our case. Schick showed that the latter choice of function ϕ leads to estimator $\widehat{\beta}_n(\phi)$ with the smallest asymptotic variance $V(\phi)$. If we consider RCA(1) model where the error process and the random coefficient process have the same variances (i. e. $\sigma^2 = \sigma_B^2$), then the smallest asymptotic variance is achieved by the choice $\phi = \frac{x}{1+x^2}$.

We proved consistency and asymptotic normality of estimator of type (2) in RCA(1) model specified in Definition 1, see [12]. We also performed a simulation study which revealed that the new estimators seem to possess better statistical properties than the conventional estimators. The aim of this paper is to derive the rate of convergence of estimators $\widehat{\beta}_n(\phi)$ in RCA(1) process for sufficiently large class of functions ϕ.

A special case $\phi(x) = x$ has been studied for instance by Basu and Roy in [1] or [2]. In the first mentioned article, the authors proved the rate of convergence of LS estimator in both univariate and multivariate RCA models. In the second paper, they proved similar results but they used general autoregressive model with fixed coefficients instead of RCA process. We can also prove this results using linear processes. Phillips and Solo in [9] showed an appealing usage of a polynomial decomposition. Since RCA model can be expressed as an infinite linear filter, we could easily derive the mentioned results using this decomposition. Unfortunately, such approach cannot be used for a general function $\phi(x)$.

We extend the results for a general class of RCA(1) estimators. First of all we have to define the class of all admissible functions $\phi(x)$. Denote $h(x) = x\phi(x)$ and presume that

A3: $h(x) > 0$ for $x \neq 0$ and $h(x)$ fulfills Lipschitz condition

$$|h(X_s) - h(X_t)| \leq c_h |X_s - X_t| \quad \text{a.s.}$$

for given process $\{X_t, t \in \mathbb{Z}\}$, all $s, t \in \mathbb{Z}$, and some constant $c_h > 0$.

Remark.

1. If process $\{X_t\}$ is uniformly bounded by some positive constant c_X then function $h(x) = x^2$ satisfies assumption A3. This can be seen by noticing that $|X_s^2 - X_t^2| = |X_s + X_t||X_s - X_t| \leq (|X_s| + |X_t|)|X_s - X_t| \leq 2c_X |X_s - X_t|$.

2. Generally, if process $\{X_t\}$ in Assumption A3 satisfies $\mathrm{P}(X_t \in I, \forall t \in \mathbb{Z}) = 1$ for some bounded interval $I \in \mathbb{R}$, then any function $h(x)$ with bounded first derivation on I satisfies assumption A3.

   This follows from the Lagrange Mean Value Theorem that states that for such function $h(x)$ and for each $x, y \in I$ there exists $z \in (x, y)$ such that $h(y) - h(x) = h'(z)(y - x)$. So $|h(X_s) - h(X_t)| = |h'(Z)| \cdot |(X_s - X_t)| \leq \max_{z \in I} |h'(z)| \cdot |(X_s - X_t)| = c_h |X_s - X_t|$ a.s..

   Thus, function $h(x) = \frac{x^2}{1+x^2}$, which corresponds to $\phi(x) = \frac{x}{1+x^2}$, fits for instance (see the discussion of the choice of ϕ above).

3. AUXILIARY LEMMAS

We employ similar techniques as Basu and Roy did in [1] where they explored the rate of convergence of LS estimator in RCA models. To prove the main result we need some auxiliary lemmas.

Lemma 1. (Accuracy of normal approximation) Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space and f and g be $\mathcal{F}$-measurable functions.

Then for any $\varepsilon > 0$

$$\sup_{x \in \mathbb{R}} \left| \mathrm{P}\left(\frac{f}{g} \leq x\right) - \Phi(x) \right| \leq \sup_{y \in \mathbb{R}} \left| \mathrm{P}(f \leq y) - \Phi(y) \right| + \mathrm{P}\Big(|g - 1| > \varepsilon\Big) + \varepsilon ,$$

where $\Phi(x)$ is the distribution function of the standard normal distribution.

P r o o f. See [7]. □

Lemma 2. (Berry-Esséen theorem) Let $\{U_t, t \in \mathbb{N}\}$ be martingale difference sequence with constant nonzero variance σ^2. Let $\mathrm{P}(|U_t| \leq c, \forall t \in \mathbb{N}) = 1$ for some constant $c > 0$.

Then there exists $d > 0$ such that for each $n \in \mathbb{N}$

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{1}{\sqrt{\sigma^2}} \cdot \frac{1}{\sqrt{n}} \sum_{t=1}^{n} U_t \leq x \right) - \Phi(x) \right| \leq d \cdot \frac{(\ln n)^3}{\sqrt{n}} .$$

P r o o f. See [6]. □

It also turned out that we need a Hoeffding-type exponential inequality for a function of RCA(1) process $\{X_t\}$. Specifically we have to ensure that

$$\mathrm{P}\left(\left| \frac{1}{n} \sum_{t=0}^{n-1} \Big(h(X_t) - \mathrm{E}h(X_t) \Big) \right| > \varepsilon \right) \leq c \cdot \mathrm{e}^{-dn\varepsilon^2} , \tag{4}$$

where $h(x)$ is given measurable function and c, d are some positive constants. This inequality is well known when $h(x) = x$ and either $\{X_t\}$ are independent or they form a martingale difference sequence. For general function $h(x)$ there were similar inequalities proved for uniformly ergodic Markov chains $\{X_t\}$ (see working paper [4]) or under some assumptions for ergodic time series (see working paper [11]). Unfortunately, none of these generalizations of Hoeffding inequality was applicable in our case so we proved inequality (4) using similar techniques as the authors of [5] did for heteroscedastic RCA(1) model. For that purpose, let us define necessary concepts and lemmas (all of them can be found in [3]).

Definition 2. Let $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ be a sequence of random vectors. Let $\mathcal{F}_{-\infty}^{t} = \sigma(\boldsymbol{V}_s, s \leq t)$, and $\mathcal{F}_{t+m}^{+\infty} = \sigma(\boldsymbol{V}_s, s \geq t+m)$ for each $t, m \in \mathbb{Z}$.

Sequence $\{\boldsymbol{V}_t\}$ is said to be α-mixing (or strong mixing) if $\lim_{m\to\infty} \alpha_m = 0$ where

$$\alpha_m = \sup_{t\in\mathbb{Z}} \left(\sup_{G\in\mathcal{F}_{-\infty}^{t}, H\in\mathcal{F}_{t+m}^{+\infty}} \Big| \mathrm{P}(G \cap H) - \mathrm{P}(G)\mathrm{P}(H) \Big| \right) .$$

Precisely, a sequence is called α-mixing of size $-a_0$ if $\alpha_m = O(m^{-a})$ for some $a > a_0$.

In order to prove the main theorem, we have to make another assumption about mutual structure between the random coefficients and the error process.

A4: $\{(B_t, Y_t), t \in \mathbb{Z}\}$ is α-mixing of size $-a$ for some $a > 2$.

Definition 3. Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of integrable random variables. Let $\mathcal{F}_t$ be a filtration.

Sequence $\{X_t, \mathcal{F}_t\}$ is called an L_p-mixingale if, for $p \geq 1$, there exist sequences of nonnegative constants $\{c_t, t \in \mathbb{Z}\}$ and $\{\zeta_m, m \in \mathbb{N}_0\}$ such that $\zeta_m \to 0$ as $m \to +\infty$, and

$$\|\mathrm{E}[X_t|\mathcal{F}_{t-m}]\|_p \leq c_t \zeta_m \ ,$$
$$\|X_t - \mathrm{E}[X_t|\mathcal{F}_{t+m}]\|_p \leq c_t \zeta_{m+1}$$

hold for all $t \in \mathbb{Z}$ and $m \in \mathbb{N}_0$. š

The L_p-mixingale is of size $-a_0 > 0$ if $\zeta_m = O(m^{-a})$ for some $a > a_0$.

Definition 4. Let $\{X_t, t \in \mathbb{Z}\}$ be a sequence of integrable random variables. Let $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ be a sequence, possibly vector-valued, let us define for each $t \in \mathbb{Z}$ filtration $\{\mathcal{F}_{t-m}^{t+m}, m \in \mathbb{N}_0\}$ such that $\mathcal{F}_{t-m}^{t+m} = \sigma(\boldsymbol{V}_{t-m}, \ldots, \boldsymbol{V}_{t+m})$.

Sequence $\{X_t\}$ is said to be near-epoch dependent in L_p-norm (L_p-NED) on $\{\boldsymbol{V}_t\}$ if, for $p > 0$, there exist sequences of nonnegative constants $\{d_t, t \in \mathbb{Z}\}$ and $\{\nu_m, m \in \mathbb{N}_0\}$ such that $\nu_m \to 0$ as $m \to +\infty$, and

$$\left\|X_t - \mathrm{E}\left[X_t|\mathcal{F}_{t-m}^{t+m}\right]\right\|_p \leq d_t \nu_m$$

hold for all $t \in \mathbb{Z}$ and $m \in \mathbb{N}_0$.

The L_p-NED is of size $-a_0 > 0$ if $\nu_m = O(m^{-a})$ for some $a > a_0$.

The idea behind using just defined concepts for proving exponential inequality (4) is briefly the following: Lipschitz function maintains L_p-NED property, L_p-NED process is L_p-mixingale under certain assumptions, and sum of L_p-mixingale items can be divided into sum of martingale difference items and some residuum. Precisely, we make use of the following lemmas:

Lemma 3. Let $\{X_t, t \in \mathbb{Z}\}$ be L_2-NED of size $-a$ on $\{\boldsymbol{V}_t\}$ with constants $\{d_t\}$. Let $h(x)$ be function that satisfies Lipschitz assumption A3 with Lipschitz constant $c_h > 0$.

Then $\{h(X_t), t \in \mathbb{Z}\}$ is also L_2-NED of size $-a$ on $\{\boldsymbol{V}_t\}$ with constants $\{c_h \cdot d_t\}$.

P r o o f. It is a special case of Theorem 17.12 in [3]. □

Lemma 4. Let $\{X_t, t \in \mathbb{Z}\}$ be an L_r-bounded zero-mean sequence, for $r > 1$. Let $\{\boldsymbol{V}_t, t \in \mathbb{Z}\}$ be α-mixing of size $-a$.

If $\{X_t\}$ is L_p-NED of size $-b$ on $\{\mathbb{V}_t\}$, for $1 \leq p < r$ with constants $\{d_t\}$, $\{X_t, \mathcal{F}_t\}$ is an L_p-mixingale of size $-\min(b, a(1/p - 1/r))$ with constants $\{c_t\}$, such that $c_t = O(\max(\|X_t\|_r, d_t))$.

P r o o f. See Theorem 17.5 in [3]. □

Lemma 5. Let $\{X_t, \mathcal{F}_t\}$ be a stationary L_1-mixingale of size -1.
There exists decomposition

$$X_t = R_t - R_{t+1} + W_t \ , \tag{5}$$

where $\mathrm{E}|R_t| < +\infty$ and $\{W_t, \mathcal{F}_t\}$ is a stationary martingale difference sequence.
Equation (5) immediately implies that

$$\sum_{t=0}^{n-1} X_t = R_0 - R_n + \sum_{t=0}^{n-1} W_t \, . \tag{6}$$

Proof. See Theorem 16.6 in [3]. □

Corollary. If sequence $\{X_t\}$ in Lemma 5 is moreover uniformly bounded L_1-mixingale of size -1 with bounded constant sequence $\{c_t, t \in \mathbb{Z}\}$ in Definition 3, then both sequences $\{R_t\}$ and $\{W_t\}$ are uniformly bounded.

Proof. We make use of proof of Theorem 16.6 stated in [3]. Let us for each $m \in \mathbb{N}$ and $t \in \mathbb{Z}$ define random variable

$$R_{m,t} = \sum_{s=0}^{m} (\mathrm{E}[X_{t+s}|\mathcal{F}_{t-1}] - X_{t-s-1} + \mathrm{E}[X_{t-s-1}|\mathcal{F}_{t-1}]) \, .$$

Now using triangular inequality and L_1-mixingale property we have that

$$\begin{aligned} |R_{m,t}| &\le \sum_{s=0}^{m} \left(\Big|\mathrm{E}[X_{t+s}|\mathcal{F}_{t-1}]\Big| + \Big|X_{t-s-1} - \mathrm{E}[X_{t-s-1}|\mathcal{F}_{t-1}]\Big| \right) \\ &\le \sum_{s=0}^{m} (c_{t+s} \cdot \zeta_{s+1} + c_{t-s-1} \cdot \zeta_{s+1}) \le 2c \cdot \sum_{s=0}^{m} \zeta_{s+1} \sim 2c \cdot \sum_{s=0}^{m} \frac{1}{m^a} = d \ , \end{aligned}$$

where $c_t \le c$ for some $c > 0$. Constant $d > 0$ exists because of $a > 1$. Thus, $\{R_{m,t}\}$ are bounded uniformly in both m and t. In the proof mentioned above there is shown that for each $t \in \mathbb{Z}$, $R_{m,t}$ converges to R_t a. s. as $m \to +\infty$. Thus, R_t are uniformly bounded in t because of boundedness of $R_{m,t}$.

Triangular inequality applied to rearranged equation (5) yields $|W_t| = |X_t - R_t + R_{t+1}| \le c_X + 2d$, which ensures uniform boundedness of W_t. □

We have to assume some additional propositions concerning RCA(1) process $\{X_t\}$ to employ the previous techniques.

A5: $\mathrm{P}(|X_0| \le c_X) = 1$ for some constant $c_X > 0$.

A6: $\mathrm{P}(|B_0| \le c_B) = 1$ for some constant $c_B > 0$.

Since RCA(1) process $\{X_t\}$ is strictly stationary, boundedness of X_0 according to A5 ensures uniform boundedness of the whole sequence $\{X_t\}$, i. e. there exists constant $c_X > 0$ such that $\mathrm{P}(|X_t| \le c_X, \forall t \in \mathbb{Z}) = 1$.

Lemma 6. Under Assumptions A5 and A6 both $\{B_t\}$ and $\{Y_t\}$ in the Definition 1 are uniformly bounded.

Proof. Firstly, sequence $\{B_t\}$ is uniformly bounded according to Assumptions A1 and A6 using its strictly stationary property.

Secondly, equation (1) gives us $|X_t| = |(\beta + B_t)X_{t-1} + Y_t| \geq |B_t X_{t-1} + Y_t| - |\beta X_{t-1}|$, consequently $|B_t X_{t-1} + Y_t| \leq |X_t| + |\beta X_{t-1}| \leq (1 + |\beta|)c_X$ a. s. owing to boundedness of $\{X_t\}$, so we know that $\{B_t X_{t-1} + Y_t\}$ is uniformly bounded.

Finally, $|B_t X_{t-1} + Y_t| \geq |Y_t| - |B_t||X_{t-1}|$, thus $|Y_t| \leq |B_t X_{t-1} + Y_t| + |B_t||X_{t-1}| \leq c$ for some constant $c > 0$ and all $t \in \mathbb{Z}$ because all processes on the right hand side of the inequality are uniformly bounded. □

Lemma 7. Let $\{X_t, t \in \mathbb{Z}\}$ be RCA(1) process that satisfies Assumptions A4 to A6. Let $\phi : \mathbb{R} \to \mathbb{R}$ be continuous measurable function such that $h(x) = x\phi(x)$ satisfies Assumption A3. Denote $Z_t = h(X_t) - \mathrm{E}h(X_t)$ for each $t \in \mathbb{Z}$.

Then there exists decomposition of process $\{Z_t, t \in \mathbb{Z}\}$ given in Lemma 5 for $\mathcal{F}_t = \sigma(B_s, Y_s; s \leq t)$ where both sequences $\{R_t\}$ and $\{W_t\}$ are uniformly bounded.

Proof. Firstly, we will show that $\{X_t\}$ is L_2-NED on $\{B_t, Y_t\}$ of arbitrary size. Let us denote $\mathcal{F}_{t-m}^{t+m} = \sigma(B_s, Y_s; s = t-m, \ldots, t+m)$ and verify Definition 4:

Definition of RCA(1) model and $\mathcal{F}_{t-m}^{t+m}$-measurability of B_s and Y_s for $s = t, t-1, \ldots, t-m$ yields

$$
\begin{aligned}
&\left\|X_t - \mathrm{E}\big[X_t|\mathcal{F}_{t-m}^{t+m}\big]\right\|_2 = \left\|(\beta + B_t)X_{t-1} + Y_t - \mathrm{E}\big[(\beta + B_t)X_{t-1} + Y_t|\mathcal{F}_{t-m}^{t+m}\big]\right\|_2 \\
&= \left\|(\beta + B_t)X_{t-1} - (\beta + B_t)\mathrm{E}\big[X_{t-1}|\mathcal{F}_{t-m}^{t+m}\big]\right\|_2 \\
&= \left\|(\beta + B_t)\left(X_{t-1} - \mathrm{E}\big[X_{t-1}|\mathcal{F}_{t-m}^{t+m}\big]\right)\right\|_2 = \ldots \\
&= \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) \cdot \left(X_{t-m-1} - \mathrm{E}\big[X_{t-m-1}|\mathcal{F}_{t-m}^{t+m}\big]\right)\right\|_2 \\
&\leq \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) X_{t-m-1}\right\|_2 + \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) \mathrm{E}\big[X_{t-m-1}|\mathcal{F}_{t-m}^{t+m}\big]\right\|_2 \\
&\leq \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) c_X\right\|_2 + \left\|\left(\prod_{i=0}^{m}(\beta + B_{t-i})\right) c_X\right\|_2 .
\end{aligned}
$$

The last inequality holds due to uniform boundedness of sequence $\{X_t\}$ by positive constant c_X. Notice that

$$
\left\|\prod_{i=0}^{m}(\beta + B_{t-i})\right\|_2 = \left(\mathrm{E}\prod_{i=0}^{m}(\beta + B_{t-i})^2\right)^{\frac{1}{2}} = \big(\mathrm{E}(\beta + B_t)^2\big)^{\frac{m+1}{2}} = \left(\beta^2 + \sigma_B^2\right)^{\frac{m+1}{2}} .
$$

So we have

$$
\left\|X_t - \mathrm{E}\big[X_t|\mathcal{F}_{t-m}^{t+m}\big]\right\|_2 \leq 2 \cdot c_X \cdot \left(\beta^2 + \sigma_B^2\right)^{\frac{m+1}{2}} .
$$

Since it is assumed that $\beta^2 + \sigma_B^2 < 1$, term $(\beta^2 + \sigma_B^2)^{\frac{m+1}{2}}$ is $O(m^{-a})$ for any $a > 0$ and thus $\{X_t\}$ is L_2-NED with constants $d_t = 2 \cdot c_X$ in Definition 4.

Secondly, $\{h(X_t)\}$ is L_2-NED on $\{B_t, Y_t\}$ of arbitrary size with constants $d_t = 2 \cdot c \cdot c_h$ (due to Lemma 3 and Lipschitz property of function $h(x)$ with some constant c_h) and so is $\{Z_t\}$ (adding constant does not violate NED condition).

Finally, if we knew that $\{Z_t\}$ is L_1-mixingale of size -1 we would have the desired decomposition of Z_t according to Lemma 5. We will show that $\{Z_t\}$ is even L_2-mixingale (and norm inequality $\|\cdot\|_1 \le \|\cdot\|_2$ will ensure its L_1-mixingale property). This can be seen by applying Lemma 4 to sequence $\{Z_t\}$ (which is L_r-bounded for any $r > 1$ because sequence $\{X_t\}$ is bounded and function $h(x)$ is continuous) and α-mixing sequence $\{B_t, Y_t\}$ of size $-a$ for some $a > 2$ (Assumption A4). $\{Z_t\}$ is L_2-NED of arbitrary size $-b$, so the size of L_2-mixingale Z_t is $-\min(b, a(1/2 - 1/r))$. Notice, that $\min(b, a(1/2 - 1/r)) = a/2 - a/r > 1$ because term a/r can be arbitrary small by increasing r. Constants c_t in Definition 3 are equal to $O(\max(\|X_t\|_r, 2 \cdot c \cdot c_h))$, they are bounded due to the boundedness of $\{X_t\}$, thus both sequences $\{R_t\}$ and $\{W_t\}$ in decomposition are uniformly bounded due to the corollary of Lemma 5. □

Lemma 8. Let $\{X_t, t \in \mathbb{Z}\}$ be RCA(1) process that satisfies Assumptions A4 to A6. Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuous measurable function such that $h(x) = x\phi(x)$ satisfies Assumption A3.

Then there exist constants $c > 0$ and $d > 0$ such that exponential inequality (4) holds.

Proof. Denote $Z_t = h(X_t) - \mathrm{E}h(X_t)$ for each $t \in \mathbb{Z}$. Then Lemma 7 gives us decomposition (6) of $\{Z_t\}$, namely $\sum_{t=0}^{n-1} Z_t = R_0 - R_n + \sum_{t=0}^{n-1} W_t$, where $\{W_t\}$ is a uniformly bounded martingale difference sequence and $\{R_t\}$ is a uniformly bounded random sequence. Now we get

$$\begin{aligned}
\mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1} Z_t\right| > \varepsilon\right) &= \mathrm{P}\left(\left|\frac{1}{n}\left(R_0 - R_n + \sum_{t=0}^{n-1} W_t\right)\right| > \varepsilon\right) \\
&\le \mathrm{P}\left(\frac{1}{n}\Big|R_0 - R_n\Big| + \left|\frac{1}{n}\sum_{t=0}^{n-1} W_t\right| > \varepsilon\right) \\
&\le \mathrm{P}\left(\Big|R_0 - R_n\Big| > n \cdot \frac{\varepsilon}{2}\right) + \mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1} W_t\right| > \frac{\varepsilon}{2}\right) \\
&\le \mathrm{P}\left(\Big|R_0\Big| + \Big|R_n\Big| > n \cdot \frac{\varepsilon}{2}\right) + c_1 \cdot \mathrm{e}^{-d_1 n \left(\frac{\varepsilon}{2}\right)^2}
\end{aligned}$$

using triangular inequality and Hoeffding inequality for bounded martingale differences (see for instance [3] Theorem 15.20). An upper bound for the first summand might be easily obtained as following:

$$\mathrm{P}\left(|R_0| + |R_n| > n \cdot \frac{\varepsilon}{2}\right) = \mathrm{P}\left(\mathrm{e}^{\varepsilon \cdot (|R_0| + |R_n|)} > \mathrm{e}^{n \cdot \frac{\varepsilon^2}{2}}\right) \le \mathrm{E}\left(\mathrm{e}^{\varepsilon \cdot (|R_0| + |R_n|)}\right) \cdot \mathrm{e}^{-n \cdot \frac{\varepsilon^2}{2}}$$

$$\leq c_2 \cdot \mathrm{e}^{-2n\left(\frac{\varepsilon}{2}\right)^2} ,$$

where finite constant c_2 can be found thanks to uniform boundedness of $\{R_t\}$. If we combine this result with the previous estimate, we get

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{t=0}^{n-1} Z_t\right| > \varepsilon\right) \leq \max(c_1, c_2) \cdot \mathrm{e}^{-\min(d_1,2)\cdot n\cdot\left(\frac{\varepsilon}{2}\right)^2} = c \cdot \mathrm{e}^{-dn\varepsilon^2}$$

which should be proved. □

4. MAIN RESULT

Now, let us formulate and prove the main theorem about the rate of convergence.

Theorem. Let $\{X_t, t \in \mathbb{Z}\}$ be RCA(1) process according to Definition 1. Let Assumptions A4 to A6 be satisfied.

Let $\phi : \mathbb{R} \to \mathbb{R}$ be a continuous measurable function such that $h(x) = x\phi(x)$ satisfies Assumption A3.

Then there exists $c > 0$ such that for each $n \in \mathbb{N}$

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\sqrt{n}\frac{\left(\widehat{\beta}_n(\phi) - \beta\right)}{\sqrt{V(\phi)}} \leq x\right) - \Phi(x)\right| \leq c \cdot \frac{(\ln n)^3}{\sqrt{n}} , \tag{7}$$

where estimator $\widehat{\beta}_n(\phi)$ is defined by (2) and asymptotic variance $V(\phi)$ is defined by (3).

Remark. This theorem covers both LS estimator (choice $\phi(x) = x$) and the estimator with the smallest asymptotic variance in model with $\sigma_B^2 = \sigma_Y^2$ (choice $\phi(x) = \frac{x}{1+x^2}$), see paragraph 2.

Proof. We basically follow similar proof from [1] for scalar RCA(1) model.

For each $n \in \mathbb{N}$ let us define

$$f_n = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} \phi(X_{t-1})(X_t - \beta X_{t-1}) ,$$

$$g_n = \frac{1}{n} \sum_{t=1}^{n} \phi(X_{t-1})X_{t-1} = \frac{1}{n} \sum_{t=1}^{n} h(X_{t-1}) .$$

Then $\sqrt{n}(\widehat{\beta}_n(\phi) - \beta) = \frac{f_n}{g_n}$. Denote $U(\phi) = \mathrm{E}\left(\phi^2(X_1)w(X_1)\right)$ where $w(x) = \sigma^2 + \sigma_B^2 x^2$ and notice that $V(\phi) = \frac{U(\phi)}{(\mathrm{E}h(X_1))^2} > 0$, so we have

$$\sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\sqrt{n}\frac{\left(\widehat{\beta}_n(\phi) - \beta\right)}{\sqrt{V(\phi)}} \leq x\right) - \Phi(x)\right| = \sup_{x\in\mathbb{R}} \left| \mathrm{P}\left(\frac{\mathrm{E}h(X_1)}{\sqrt{U(\phi)}}\frac{f_n}{g_n} \leq x\right) - \Phi(x)\right| .$$

Lemma 1 gives us for any $\varepsilon > 0$

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}\left(\frac{\mathrm{E}h(X_1)}{\sqrt{U(\phi)}}\frac{f_n}{g_n}\le x\right)-\Phi(x)\right|\le\sup_{y\in\mathbb{R}}\left|\mathrm{P}\left(\frac{f_n}{\sqrt{U(\phi)}}\le y\right)-\Phi(y)\right| + \mathrm{P}\left(\left|\frac{g_n}{\mathrm{E}h(X_1)}-1\right|>\varepsilon\right)+\varepsilon\,. \qquad (8)$$

It can be easily derived that $\{\phi(X_{t-1})(X_t-\beta X_{t-1})\}$ is $\sigma\,(B_s,Y_s;s\le t)$-martingale difference sequence with variance $U(\phi)$ defined previously. Function ϕ is continuous, process $\{X_t\}$ is uniformly bounded, thus $\{\phi(X_{t-1})(X_t-\beta X_{t-1})\}$ is also uniformly bounded. So Lemma 2 can be applied to the first term on the right hand side of inequality (8) and we have

$$\sup_{y\in\mathbb{R}}\left|\mathrm{P}\left(\frac{f_n}{\sqrt{U(\phi)}}\le y\right)-\Phi(y)\right|\le d\cdot\frac{(\ln n)^3}{\sqrt{n}}\,,$$

where $d>0$ is some constant.

The second term on the right hand side of inequality (8) can be arranged into

$$\mathrm{P}\left(\left|\frac{g_n}{\mathrm{E}h(X_1)}-1\right|>\varepsilon\right)=\mathrm{P}\Big(\left|g_n-\mathrm{E}h(X_1)\right|>\varepsilon\cdot\mathrm{E}h(X_1)\Big) = \mathrm{P}\left(\left|\frac{1}{n}\sum_{t=1}^{n}\Big(h(X_{t-1})-\mathrm{E}h(X_{t-1})\Big)\right|>\varepsilon\cdot\mathrm{E}h(X_1)\right)\,,$$

using definition of g_n and strict stationarity of $\{X_t\}$. All conditions of Lemma 8 are met and we have

$$\mathrm{P}\left(\left|\frac{1}{n}\sum_{t=1}^{n}\Big(h(X_{t-1})-\mathrm{E}h(X_{t-1})\Big)\right|>\varepsilon\cdot\mathrm{E}h(X_1)\right)\le p\cdot\mathrm{e}^{-qn\varepsilon^2}\,,$$

where $p,q>0$ are some constants.

If we sum up all derived results we gain that for any $\varepsilon>0$ there exist positive constants d,p,q such that

$$\sup_{x\in\mathbb{R}}\left|\mathrm{P}\Big(\sqrt{n}\,\Big(\widehat{\beta}_n(\phi)-\beta\Big)\le x\Big)-G(x)\right|\le d\cdot\frac{(\ln n)^3}{\sqrt{n}}+p\cdot\mathrm{e}^{-qn\varepsilon^2}+\varepsilon\,.$$

Setting $\varepsilon=\frac{(\ln n)^3}{\sqrt{n}}$ we obtain the desired upper bound $c\cdot\frac{(\ln n)^3}{\sqrt{n}}$ for some positive constant c. □

ACKNOWLEDGEMENT

This work was partially supported by Grant Agency of the Czech Republic under Grant 201/05/H007.

(Received November 24, 2005.)

REFERENCES

[1] 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.

[2] A. K. Basu and S. Sen Roy: On rates of convergence in the central limit theorem for parameter estimation in random coefficient autoregressive models. J. Indian Statist. Assoc. *26* (1988), 19–25.

[3] J. Davidson: Stochastic Limit Theory. (Advanced Texts in Econometrics.) Oxford University Press, New York, Reprinted 2002.

[4] P. W. Glynn and D. Ormoneit: Hoeffding's Inequality for Uniformly Ergodic Markov Chains. Working paper, http://ai.stanford.edu/ ormoneit/publications/hoeffding.ps

[5] H. Janečková and Z. Prášková: CWLS and ML estimates in a heteroscedastic RCA(1) model. Statist. Decisions *22* (2004), 245–259.

[6] Y. Kato: Rates of convergence in central limit theorem for martingale differences. Bull. Math. Statist *18* (1979), 1–8.

[7] R. Michel and J. Pfanzagl: The accuracy of the normal approximation for minimum contrast estimate. Z. Wahrsch. Verw. Gebiete *18* (1971), 73–84.

[8] D. F. Nicholls and B. G. Quinn: Random Coefficient Autoregressive Models: An Introduction. (Lecture Notes in Statistics 11.) Springer, New York 1982.

[9] P. C. B. Phillips and V. Solo: Asymptotics for linear processes. Ann. Statist. *20* (1992), 971–1001.

[10] A. Schick: $\sqrt{n}$-consistent estimation in a random coefficient autoregressive model. Austral. J. Statist. *38* (1996), 155–160.

[11] Y. Tang: A Hoeffding-type Inequality for Ergodic Time Series. Working paper, http://www4.stat.ncsu.edu/ sghosal/papers/Tang.pdf

[12] P. Vaněček: Estimators of generalized RCA models. In: Proc. WDS'04 *Part I* (2004), pp. 35–40.

Pavel Vaněček, Department of Probability and Mathematical Statistics, Faculty of Mathematics and Physics – Charles University, Sokolovská 83, 186 75 Praha 8. Czech Republic.

e-mail: vanecek@karlin.mff.cuni.cz