Zhao et al. *Journal of Inequalities and Applications* (2018) 2018:82
https://doi.org/10.1186/s13660-018-1680-4

Journal of Inequalities and Applications
a SpringerOpen Journal

RESEARCH Open Access

Variable selection in generalized random coefficient autoregressive models

Zhiwen Zhao[1*], Yangping Liu[1] and Cuixin Peng[2]

*Correspondence: georgejilin@126.com
[1]College of Mathematics, Jilin Normal University, Siping, P.R. China
Full list of author information is available at the end of the article

Abstract

In this paper, we consider the variable selection problem of the generalized random coefficient autoregressive model (GRCA). Instead of parametric likelihood, we use non-parametric empirical likelihood in the information theoretic approach. We propose an empirical likelihood-based Akaike information criterion (AIC) and a Bayesian information criterion (BIC).

MSC: Primary 62M10; secondary 91B62

Keywords: Empirical likelihood; Akaike information criterion; Bayesian information criterion; Generalized random coefficient autoregressive model; Variable selection

1 Introduction

Consider the following p-order generalized random coefficient autoregressive model:

$$Y_t = \Phi_t^\tau Y(t-1) + \varepsilon_t, \tag{1}$$

where τ denotes the transpose of a matrix or vector, $\Phi_t = (\Phi_{t1}, \dots, \Phi_{tp})^\tau$ is a random coefficient vector, $Y(t-1) = (Y_{t-1}, \dots, Y_{t-p})^\tau$, and $\{\binom{\Phi_t}{\varepsilon_t}, t = 0, \pm 1, \pm 2, \dots\}$ is a sequence of i.i.d. random vectors with $E(\Phi_t) = \phi = (\phi_1, \dots, \phi_p)$, $E(\varepsilon_t) = 0$, and $\operatorname{Var}\binom{\Phi_t}{\varepsilon_t} = \begin{pmatrix} V_\phi & \sigma_{\Phi\varepsilon} \\ \sigma_{\Phi\varepsilon}^\tau & \sigma_\varepsilon^2 \end{pmatrix}$.

As a generalization of the usual autoregressive model, the random coefficient autoregressive (RCAR) model (*cf.* [1, 2]), the Markovian bilinear model and its generalization, and the random coefficient exponential autoregressive model (*cf.* [3–5]), model (1) was first introduced by Hwang and Basawa [6]. GRCA has become one of the important models in the nonlinear time series context. In recent years, GRCA has been studied by many authors. For instance, Hwang and Basawa [7] established the local asymptotic normality of a class of generalized random coefficient autoregressive processes. Carrasco and Chen [8] provided the tractable sufficient conditions that simultaneously imply strict stationarity, finiteness of higher-order moments, and β-mixing with geometric decay rates. Zhao and Wang [9] constructed confidence regions for the parameters of model (1) by using an empirical likelihood method. Furthermore, Zhao et al. [10] also considered the problem of testing the constancy of the coefficients in the stationary one-order generalized random coefficient autoregressive model. In this paper, we consider the variable selection problem of the GRCA based on the empirical likelihood method.

Many model selection procedures have been proposed in the statistical literature, including the adjusted R^2 (see Theil [11]), the AIC (see Akaike [12]), BIC (see Schwarz [13]), Mallow'S C_p (see Mallows [14]). Other criteria in the literature include Hannan and Quinn's criterion [15], Geweke and Meese's criterion [16], Cavanaugh's Kullback information criterion [17], and the deviance information criterion of Spiegelhalter et al. [18]. Also, Tsay [19], Hurvich and Tsai [20] and Pötscher [21] have studied model selection methods in time series models. Recently, the model selection problem has been extended to moment selection as in Andrews [22], Andrews and Lu [23] and Hong et al. [24]. These model selection methods are concerned with parsimony, as was stressed in Zellner et al. [25], as well as accuracy or power in choosing models.

In this paper, we develop an information theoretic approach to variable selection problem of GRCA. Specifically, instead of parametric likelihood, we use non-parametric empirical likelihood (see Owen [26, 27]) in the information theoretic approach. We propose an empirical likelihood-based Akaike information criterion (EAIC) and a Bayesian information criterion (EBIC).

The paper proceeds as follows. The next section is concerned with the methodology and the main results. Section 3 is devoted to the proofs of the main results.

Throughout the paper, we use the symbols "$\xrightarrow{d}$" and "$\xrightarrow{p}$" to denote convergence in distribution and convergence in probability, respectively. We abbreviate "almost surely" and "independent identical distributed" to "a.s." and "i.i.d.", respectively. $o_p(1)$ means a term which converges to zero in probability. $O_p(1)$ means a term which is bounded in probability. Furthermore, the Kronecker product of the matrices A and B is denoted by $A \otimes B$, and $\|M\|$ denotes the L_2 norm for vector or matrix M.

2 Methods and main results

In this section, we will first propose the empirical likelihood-based information criteria for choice of a GRCA, then we investigate the asymptotic properties of the new variable selection method.

2.1 Empirical likelihood-based information criteria

Hwang and Basawa [6] derived the conditional least-squares estimator $\hat{\phi}$ of ϕ, which is given by

$$\hat{\phi} = \left(\sum_{t=1}^{n} Y(t-1)Y^{\tau}(t-1) \right)^{-1} \left(\sum_{t=1}^{n} Y_t Y(t-1) \right).$$

By using the estimating equation of the conditional least-squares estimator, we can obtain the following score function:

$$\sum_{t=1}^{n} \bigl(Y_t Y(t-1) - Y(t-1)Y^{\tau}(t-1)\phi \bigr) = \sum_{t=1}^{n} G_t(\phi),$$

where $G_t(\phi) = Y_t Y(t-1) - Y(t-1)Y^{\tau}(t-1)\phi$. Following Owen [26], the empirical likelihood statistic for ϕ is defined as

$$\tilde{l}(\phi) = -2 \max_{\sum_{t=1}^{n} p_t G_t(\phi)=0} \sum_{t=1}^{n} \log(np_t),$$

where $p_1, \ldots, p_n$ are all sets of nonnegative numbers summing to 1. By using the Lagrange multiplier method, let

$$G = \sum_{t=1}^{n} \log(np_t) - n\lambda^\tau \sum_{t=1}^{n} p_t G_t(\phi) + \gamma \left(\sum_{t=1}^{n} p_t - 1 \right).$$

After simple algebraic calculation, we have

$$\frac{\partial G}{\partial p_t} = \frac{1}{p_t} - n\lambda^\tau G_t(\phi) + \gamma, \quad t = 1, \ldots, n.$$

Note that $\sum_{t=1}^{n} p_t = 1$ and $\sum_{t=1}^{n} p_t G_t(\phi) = 0$. So we have $\gamma = -n$ and $p_t = \frac{1}{n(1+\lambda^\tau G_t(\phi))}$, which implies that

$$\tilde{l}(\phi) = 2\sum_{t=1}^{n} \log\bigl(1 + \lambda^\tau G_t(\phi)\bigr), \tag{2}$$

where λ is the solution of the equation

$$\frac{1}{n}\sum_{t=1}^{n} \frac{G_t(\phi)}{1 + \lambda^\tau G_t(\phi)} = 0. \tag{3}$$

The definition of $\tilde{l}(\phi)$ relies on finding a positive $p_t's$ such that $\sum_{t=1}^{n} p_t G_t(\phi) = 0$ for each ϕ. The solution exists if and only if the convex hull of the $G_t(\phi)$, $t = 1, 2, \ldots, n$ contains zero as an inner point. When the model is correct, the solution exists with probability tending to 1 as the sample size $n \to \infty$ for ϕ in a neighborhood of ϕ_0. However, for finite n and at some ϕ value, the equation often does not have a solution in p_t. To avoid this problem, we introduce the adjusted empirical likelihood.

Further let $\bar{G}_n = n^{-1}\sum_{t=1}^{n} p_t G_t(\phi)$ and define $G_{n+1} = -a_n \bar{G}_n$ for some positive constant a_n. We adjust the profile empirical log-likelihood ratio function to

$$\begin{aligned} l(\phi) &= -2 \max_{\sum_{t=1}^{n+1} p_t G_t(\phi) = 0} \sum_{t=1}^{n+1} \log\bigl((n+1)p_t\bigr) \\ &= 2\sum_{t=1}^{n+1} \log\bigl\{1 + \tilde{\lambda}^\tau G_t(\phi)\bigr\} \end{aligned} \tag{4}$$

with $\tilde{\lambda} = \tilde{\lambda}(\phi)$ being the solution of

$$\frac{1}{n+1}\sum_{t=1}^{n+1} \frac{G_t(\phi)}{1 + \lambda^\tau G_t(\phi)} = 0. \tag{5}$$

Since 0 always lies on the line connecting $\bar{G}_n$ and G_{n+1}, the adjusted empirical log-likelihood ratio function is well defined after adding a pseudo-value G_{n+1} to the data set. The adjustment is particularly useful so that a numerical program does not crash simply because some undesirable ϕ is assessed.

A full GRCA assumes that y_t relates to $\Phi_t^\tau Y(t-1)$ with $E(\Phi_t) = \phi$ being unknown parameter of size p. Let s be a subset of $\{1, 2, \dots, p\}$, and $Y^{[s]}(t-1)$ and $\phi^{[s]}$ be subvectors of $Y(t-1)$ and ϕ containing entries in positions specified by s. Consider the pth-order GRCA specified by $E(G_t(\phi)) = 0$ and a submodel specified by $E(G_t^{[s]}(\phi^{[s]})) = 0$, where $G_t^{[s]}(\phi^{[s]}) = Y_t Y^{[s]}(t-1) - Y^{[s]}(t-1)(Y^{[s]}(t-1))^\tau \phi^{[s]}$. For a given s, let $G_t^{[s]} = Y_t Y^{[s]}(t-1) - Y^{[s]}(t-1)(Y^{[s]}(t-1))^\tau \phi^{[s]}$, $\bar{G}_n^{[s]} = n^{-1} \sum_{t=1}^n G_t^{[s]}$ and $G_{n+1}^{[s]} = -a_n \bar{G}_n^{[s]}$ for some positive constant a_n. The adjusted empirical log-likelihood ratio becomes

$$\begin{aligned} l(\phi^{[s]}) &= -2 \max_{\sum_{t=1}^{n+1} p_t G_t^{[s]} = 0} \sum_{t=1}^{n+1} \log\bigl((n+1)p_t\bigr) \\ &= 2 \sum_{t=1}^{n+1} \log\bigl\{1 + \tilde{\lambda}^\tau G_t^{[s]}\bigr\} \end{aligned} \tag{6}$$

with $\tilde{\lambda} = \tilde{\lambda}(\phi)$ being the solution of

$$\frac{1}{n+1} \sum_{t=1}^{n+1} \frac{G_t^{[s]}}{1 + \lambda^\tau G_t^{[s]}} = 0. \tag{7}$$

We define the adjusted profile empirical log-likelihood ratio as

$$l(s) = \inf\bigl\{l(\phi^{[s]}) : \phi^{[s]}\bigr\}. \tag{8}$$

The empirical likelihood versions of AIC and BIC are then defined as

$$\mathrm{EAIC} = l(s) + 2k, \tag{9}$$

$$\mathrm{EBIC} = l(s) + k\log(n), \tag{10}$$

where k is the cardinality of s.

After $l(s)$ is evaluated for all s, we select the model with the minimum EAIC or EBIC value.

2.2 Asymptotic properties

It is well known that under some mild conditions the parametric BIC is consistent for variable selection while the parametric AIC is not. Similarly, we can prove that, when p is constant, EBIC is consistent but EAIC is not.

For purposes of illustration, in what follows, we rewrite the model in the following matrix form (see Hwang and Basawa [6]): let $U_t = (\varepsilon_t, 0, 0, \dots, 0)^\tau$ are $p \times 1$ vectors, $\tilde{\Phi}_{tj} = \Phi_{tj} - \phi_j$, $j = 1, \dots, p$,

$$B = \begin{pmatrix} \phi_1 & \phi_2 & \cdots & \phi_{p-1} & \phi_p \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & 0 \end{pmatrix}_{p\times p}, \qquad C_t = \begin{pmatrix} \tilde{\Phi}_{t1} & \tilde{\Phi}_{t2} & \cdots & \tilde{\Phi}_{tp} \\ 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}_{p\times p}.$$

Then model (1) can be written as

$$Y(t) = (B + C_t)Y(t-1) + U_t. \tag{11}$$

In order to obtain our theorems, we need the following regularity conditions:
($\mathbf{A_1}$) All the eigenvalues of the matrix $E(C_t \otimes C_t) + (B \otimes B)$ are less than unity in modulus.
($\mathbf{A_2}$) $EY_t^6 < \infty$.

Remark 1 As for the condition ($\mathbf{A_1}$) and the sufficient condition for $E|y_t|^{2m} < \infty$ ($m = 1, 2, \ldots$), we refer to Hwang and Basawa [6].

Theorem 2.1 *Let $A = E(G_t(\phi_0)G_t^\tau(\phi_0))$ and $B = E((\partial G_t(\phi)/\partial\phi)|_{\phi=\phi_0})$. If ($\mathbf{A_1}$) and ($\mathbf{A_2}$) hold, then there exists a sequence of adjusted empirical likelihood estimates $\tilde{\phi}$ of ϕ such that*

$$\sqrt{n}(\tilde{\phi} - \phi) \rightarrow N\bigl(0, \bigl(B^\tau A^{-1}B\bigr)^{-1}\bigr) \tag{12}$$

and

$$\sqrt{n}(\tilde{\lambda} - \lambda) \rightarrow N(0, U), \tag{13}$$

where $U = A^{-1} - A^{-1}B(B^\tau A^{-1}B)^{-1}B^\tau A^{-1}$.

Note that when a submodel s is a true model, it implies $\phi_0^{[\bar{s}]} = 0$. That is, components of ϕ_0 not in s are zero. Therefore, Y_t only relates to the variables in positions specified by s. The following theorem shows that when $\phi_0^{[\bar{s}]} = 0$ is true, then adjusted empirical log-likelihood ratio statistic has a chi-squared limiting distribution with k fewer degrees of freedom.

Theorem 2.2 *Assume that ($\mathbf{A_1}$) and ($\mathbf{A_2}$) hold and $\phi_0^{[\bar{s}]} = 0$ for a submodel s of size k. Then when $a_n = o_p(n^{\frac{1}{2}})$, we have $l(s) \rightarrow \chi^2_{p-k}$ in distribution as $n \rightarrow \infty$.*

When the null hypothesis of $\phi_0^{[\bar{s}]} = 0$ is not true, the likelihood ratio go to ∞ as $n \rightarrow \infty$. We state the following theorem in terms of the adjusted empirical likelihood which also applies to the usual empirical likelihood.

Theorem 2.3 *Assume that ($\mathbf{A_1}$) and ($\mathbf{A_2}$) hold and $a_n = o_p(n^{\frac{1}{2}})$. Then for any $\phi \neq \phi_0$ such that $E(G_t(\phi)) \neq 0$, $l(s) \rightarrow \infty$ in probability as $n \rightarrow \infty$.*

The following theorem indicates that, when p is constant, EBIC is consistent but EAIC is not.

Theorem 2.4 *Assume that ($\mathbf{A_1}$) and ($\mathbf{A_2}$) hold and if there exists a subset s_0 of $1, 2, \ldots, p$ such that, for any other subset s, $E(G_t^{[s]}(\phi^{[s]})) = 0$ for some ϕ if and only if s contains s_0. Then, EBIC is consistent and EAIC is not consistent.*

3 Proofs of the main results

In order to prove Theorem 2.1, we first present several lemmas.

Lemma 3.1 *Assume that* $(\mathbf{A_1})$ *and* $(\mathbf{A_2})$ *hold. Then A is positive definite and B has rank p.*

Proof After simple algebra calculation, we have, for any nonzero vector $c = (c_1, \ldots, c_p) \in R^P$,

$$c^\tau Ac = E\bigl(c^\tau G_t(\phi)G_t^\tau(\phi)c\bigr) = E\bigl(\bigl(c^\tau Y(t-1)\bigr)^2 \operatorname{Var}\bigl(Y_t|Y(t-1)\bigr)\bigr).$$

Note that the conditional distribution of Y_t, given $Y(t-1)$, is not a degenerate distribution, which implies that $\operatorname{Var}(Y_t|Y(t-1)) > 0$ a.s. It follows that $(c^\tau Y(t-1))^2 \operatorname{Var}(Y_t|Y(t-1)) \geq 0$ a.s. Therefore, $c^\tau Ac = 0$ if and only if $c^\tau Y(t-1) = 0$ a.s. Without loss of generality, suppose that the first component c_1 of c is 1, so $Y_{t-1} = -c_2Y_{t-2} - \cdots - c_pY_{t-p}$, which is contradictory with the fact that the conditional distribution of Y_{t-1}, given $(Y_{t-2}, \ldots, Y_{t-p})$, is not degenerate. Hence $c^\tau Ac > 0$. That is, A is positive definite.

Similarly, we can also prove that B has rank p. The proof of Lemma 3.1 is thus complete. □

Lemma 3.2 *Assume that* $(\mathbf{A_1})$ *and* $(\mathbf{A_2})$ *hold. Then when $a_n = o(n^{\frac{1}{2}})$, we have*

$$\sup_\phi \left\| \frac{1}{n+1} \sum_{t=1}^{n+1} G_t(\phi)G_t^\tau(\phi) \right\| = O(1) \quad (a.s.), \tag{14}$$

uniformly about $\phi \in \{\phi | \|\phi - \phi_0\| \leq n^{-\frac{1}{3}}\}$.

Proof Note that

$$\begin{aligned}
&\sup_\phi \left\| \frac{1}{n+1} \sum_{t=1}^{n+1} G_t(\phi)G_t^\tau(\phi) \right\| \\
&\quad\leq \sup_\phi \left\| \frac{1}{n+1} \sum_{t=1}^{n} G_t(\phi)G_t^\tau(\phi) \right\| + \sup_\phi \frac{1}{n+1} a_n^2 \left\| \frac{1}{n} \sum_{t=1}^{n} G_t(\phi) \right\|^2 \\
&\quad\leq \sup_\phi \left\| \frac{1}{n+1} \sum_{t=1}^{n} G_t(\phi)G_t^\tau(\phi) - \frac{1}{n+1} \sum_{t=1}^{n} G_t(\phi_0)G_t^\tau(\phi_0) \right\| \\
&\qquad+ \left\| \frac{1}{n+1} \sum_{t=1}^{n} G_t(\phi_0)G_t^\tau(\phi_0) \right\| + \sup_\phi \frac{1}{n+1} a_n^2 \left\| \frac{1}{n} \sum_{t=1}^{n} G_t(\phi) \right\|^2 \\
&\quad\triangleq L_{n1} + L_{n2} + L_{n3}.
\end{aligned} \tag{15}$$

First, note that

$$\begin{aligned}
L_{n1} &= \sup_\phi \left\| \frac{1}{n+1} \sum_{t=1}^{n} \bigl(Y(t-1)Y^\tau(t-1)\bigl(Y^\tau(t-1)(\phi - \phi_0)\bigr)\bigr) \right\| \\
&\leq \frac{1}{n+1} \sum_{t=1}^{n} \bigl\| Y(t-1) \bigr\|^3 \sup_\phi \|\phi - \phi_0\|.
\end{aligned}$$

By the ergodic theorem, we have

$$\frac{1}{n+1}\sum_{t=1}^{n}\|Y(t-1)\|^{3}=O(1)\quad (a.s.). \tag{16}$$

Further, note that

$$\sup_{\phi}\|\phi-\phi_0\|=O(1). \tag{17}$$

This, together with (16), proves that

$$L_{n1}=O\bigl(n^{-\frac{1}{3}}\bigr)\quad (a.s.). \tag{18}$$

Again by the ergodic theorem, we can prove that

$$L_{n2}=O(1)\quad (a.s.). \tag{19}$$

Finally, we prove that

$$L_{n3}=O\bigl(n^{-\frac{1}{3}}\bigr)\quad (a.s.). \tag{20}$$

Note that

$$\sup_{\phi}\left\|\frac{1}{n}\sum_{t=1}^{n}G_t(\phi)\right\|\le\sup_{\phi}\left\|\frac{1}{n}\sum_{t=1}^{n}\bigl(G_t(\phi)-G_t^{\tau}(\phi_0)\bigr)\right\|+\left\|\frac{1}{n}\sum_{t=1}^{n}G_t(\phi_0)\right\|.$$

Similar to the proof of (18), we can show that

$$\sup_{\phi}\left\|\frac{1}{n}\sum_{t=1}^{n}\bigl(G_t(\phi)-G_t^{\tau}(\phi_0)\bigr)\right\|=O\bigl(n^{-\frac{1}{3}}\bigr)\quad (a.s.). \tag{21}$$

In what follows, we consider $\|\frac{1}{n}\sum_{t=1}^{n}G_t(\phi_0)\|$.

Denote the ith component of $G_t(\phi_0)$ by $G_{ti}(\phi_0)$. Then $\{G_{ti}(\phi_0), 1\le i\le p\}$ is a stationary ergodic martingale difference sequence with $E(G_{ti}(\phi_0))=0$ and $E((G_{ti}(\phi_0))^2)<\infty$. By the law of the iterated logarithm of martingale difference sequence, we have, for $1\le i\le p$,

$$\frac{1}{n}\sum_{t=1}^{n}G_{ti}^{\tau}(\phi_0)=O\bigl(n^{-\frac{1}{2}}\bigl(\log_2^n\bigr)^{\frac{1}{2}}\bigr)\quad (a.s.).$$

It follows that

$$\frac{1}{n}\sum_{t=1}^{n}G_{t}^{\tau}(\phi_0)=O\bigl(n^{-\frac{1}{2}}\bigl(\log_2^n\bigr)^{\frac{1}{2}}\bigr)\quad (a.s.). \tag{22}$$

Then, by (21) and (22), we have

$$\sup_{\phi}\left\|\frac{1}{n}\sum_{t=1}^{n}G_t(\phi)\right\|=O\bigl(n^{-\frac{1}{3}}\bigr)\quad (a.s.). \tag{23}$$

Therefore

$$
\begin{aligned}
L_{n3} &= O\bigl(n^{-1}\bigr)o\bigl(n^{\frac{1}{2}}\bigr)o\bigl(n^{\frac{1}{2}}\bigr)O\bigl(n^{-\frac{1}{3}}\bigr)O\bigl(n^{-\frac{1}{3}}\bigr) \quad (a.s.) \\
&= o\bigl(n^{-\frac{2}{3}}\bigr) \quad (a.s.).
\end{aligned} \tag{24}
$$

This, together with (18) and (19), proves Lemma 3.2. □

Lemma 3.3 *Assume that* $(\mathbf{A_1})$ *and* $(\mathbf{A_2})$ *hold. Then when* $a_n = o(n^{\frac{1}{2}})$, *we have*

$$
\max_{1\le t\le n+1} \sup_{\phi} \bigl\| G_t(\phi) \bigr\| = o\bigl(n^{\frac{1}{3}}\bigr) \quad (a.s.), \tag{25}
$$

uniformly about $\phi \in \{\phi \mid \|\phi - \phi_0\| \le n^{-\frac{1}{3}}\}$.

Proof Note that

$$
\begin{aligned}
\max_{1\le t\le n+1} \sup_{\phi} \bigl\| G_t(\phi) \bigr\| &\le \max_{1\le t\le n} \sup_{\phi} \bigl\| G_t(\phi) \bigr\| + \sup_{\phi} \left\| a_n \frac{1}{n} \sum_{t=1}^{n} G_t(\phi) \right\| \\
&\triangleq K_{n1} + K_{n2}.
\end{aligned}
$$

From (23), together with $a_n = o(n^{\frac{1}{2}})$, it follows immediately that

$$
K_{n2} = o\bigl(n^{\frac{1}{3}}\bigr) \quad (a.s.). \tag{26}
$$

The next step in the proof is to show that

$$
K_{n1} = o\bigl(n^{\frac{1}{3}}\bigr) \quad (a.s.). \tag{27}
$$

By the Fubini theorem, we have, for any positive integer k,

$$
\begin{aligned}
\infty &> E\Bigl(\sup_{\phi} \bigl\| G_t(\phi) \bigr\|\Bigr)^3 \\
&= \int_0^{\infty} P\Bigl(\Bigl(\sup_{\phi} \bigl\| G_t(\phi) \bigr\|\Bigr)^3 > s\Bigr)\, ds \\
&= \sum_{n=1}^{\infty} \int_{(n-1)k^3}^{nk^3} P\Bigl(\Bigl(\sup_{\phi} \bigl\| G_t(\phi) \bigr\|\Bigr)^3 > s\Bigr)\, ds \\
&\ge \sum_{n=1}^{\infty} \int_{(n-1)k^3}^{nk^3} P\Bigl(\Bigl(\sup_{\phi} \bigl\| G_t(\phi) \bigr\|\Bigr)^3 > nk^3\Bigr)\, ds \\
&= \sum_{n=1}^{\infty} P\Bigl(\Bigl(\sup_{\phi} \bigl\| G_t(\phi) \bigr\|\Bigr)^3 > nk^3\Bigr) k^3\, ds.
\end{aligned}
$$

Thus, using the ergodic theorem,

$$
\sum_{n=1}^{\infty} P\Bigl(\sup_{\phi} \bigl\| G_n(\phi) \bigr\| > n^{\frac{1}{3}} k\Bigr) < \infty. \tag{28}
$$

By the Borel–Cantelli lemma, we know that

$$P\Bigl(\sup_{\phi}\bigl\|G_n(\phi)\bigr\| > n^{\frac{1}{3}}k \text{ i.o.}\Bigr) = 0, \tag{29}$$

so that

$$\sup_{\phi}\bigl\|G_n(\phi)\bigr\| \le n^{\frac{1}{3}}k \quad (a.s.). \tag{30}$$

Take $k = \frac{1}{m}$, then there exists Q_m with $P(Q_m) = 0$, such that, for any $\omega \in Q_m^c$,

$$\frac{\sup_{\phi}\|G_n(\phi)\|}{n^{\frac{1}{3}}} \le \frac{1}{m}. \tag{31}$$

Further, let $Q = \bigcup_{m=1}^{\infty} Q_m$. Then

$$\lim_{n\to\infty} \frac{\sup_{\phi}\|G_n(\phi)\|}{n^{\frac{1}{3}}} = 0, \tag{32}$$

which, together with the fact that $P(Q) = 0$, implies that

$$\max_{1\le t\le n}\sup_{\phi}\bigl\|G_t(\phi)\bigr\| = o\bigl(n^{\frac{1}{3}}\bigr) \quad (a.s.). \tag{33}$$

The proof is complete. □

Lemma 3.4 *Assume that* $(\mathbf{A_1})$ *and* $(\mathbf{A_2})$ *hold. Then when* $a_n = o(n^{\frac{1}{2}})$, *we have*

$$\sup_{\phi}\bigl\|\lambda(\phi)\bigr\| = O\bigl(n^{-\frac{1}{3}}\bigr) \quad (a.s.), \tag{34}$$

uniformly about $\phi \in \{\phi \mid \|\phi - \phi_0\| \le n^{-\frac{1}{3}}\}$.

Proof Write $\|\lambda(\phi)\| = \rho(\phi)\theta(\phi)$, where $\rho(\phi) > 0$ and $\|\theta(\phi)\| = 1$. Further let

$$Q_{1,n+1}(\phi,\lambda) = \frac{1}{n+1}\sum_{t=1}^{n+1}\frac{G_t(\phi)}{1+\lambda^{\tau}(\phi)G_t(\phi)}. \tag{35}$$

Then

$$\begin{aligned}
0 &= \bigl\|Q_{1,n+1}(\phi,\lambda)\bigr\| \\
&\ge \left|\frac{1}{n+1}\sum_{t=1}^{n+1}\frac{\theta^{\tau}(\phi)G_t(\phi)}{1+\lambda^{\tau}(\phi)G_t(\phi)}\right| \\
&\ge \left|\frac{1}{n+1}\rho(\phi)\sum_{t=1}^{n+1}\frac{\theta^{\tau}(\phi)G_t(\phi)G_t^{\tau}(\phi)\theta(\phi)}{1+\rho(\phi)\theta^{\tau}(\phi)G_t(\phi)}\right| - \left|\frac{1}{n+1}\sum_{t=1}^{n+1}\theta^{\tau}(\phi)G_t(\phi)\right| \\
&\ge \frac{\rho(\phi)\theta^{\tau}(\phi)(\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)G_t^{\tau}(\phi))\theta(\phi)}{\max_{1\le t\le n}\{1+\rho(\phi)\theta^{\tau}(\phi)G_t(\phi)\}} - \left|\frac{1}{n+1}\sum_{t=1}^{n+1}\theta^{\tau}(\phi)G_t(\phi)\right|,
\end{aligned}$$

which implies that

$$\begin{aligned}\frac{\rho(\phi)\theta^{\tau}(\phi)(\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)G_t^{\tau}(\phi))\theta(\phi)}{\max_{1\leq t\leq n}\{1+\rho(\phi)\theta^{\tau}(\phi)G_t(\phi)\}} &\leq \left|\frac{1}{n+1}\sum_{t=1}^{n+1}\theta^{\tau}(\phi)G_t(\phi)\right| \\ &\leq \left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)\right\|. \end{aligned}\tag{36}$$

Further, by the ergodic theorem, we have

$$\left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi_0)G_t^{\tau}(\phi_0)-A\right\| = o(1) \quad (a.s.),\tag{37}$$

where $A = E(G_t(\phi_0)G_t^{\tau}(\phi_0))$.

Since

$$\begin{aligned}0 &\leq \sup_{\phi}\left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)G_t^{\tau}(\phi)-A\right\| \\ &\leq \sup_{\phi}\left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)G_t^{\tau}(\phi)-\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi_0)G_t^{\tau}(\phi_0)\right\| \\ &\quad+\left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi_0)G_t^{\tau}(\phi_0)-A\right\|,\end{aligned}$$

we have from (18) and (37)

$$\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)G_t^{\tau}(\phi) = A+o(1) \quad (a.s.),\tag{38}$$

which implies that

$$\theta^{\tau}(\phi)\left(\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)G_t^{\tau}(\phi)\right)\theta(\phi) \geq \sigma_{\min}+o(1) \quad (a.s.),\tag{39}$$

where $\sigma_{\min}$ is the smallest eigenvalue and the largest eigenvalue of A. This, together with Lemma 3.1 and (36), proves that

$$\begin{aligned}&\sup_{\phi}\left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)\right\| \\ &\quad\geq \sup_{\phi}\rho(\phi)\left(\sigma_{\min}+o(1)-\left(\max_{1\leq t\leq n+1}\sup_{\phi}\left\|G_t(\phi)\right\|\right)\left(\sup_{\phi}\left\|\frac{1}{n+1}\sum_{t=1}^{n+1}G_t(\phi)\right\|\right)\right).\end{aligned}$$

Combined with (23) and Lemma 3.3, this establish (34) and completes the proof. □

Lemma 3.5 *Assume that* ($\mathbf{A_1}$) *and* ($\mathbf{A_2}$) *hold, and* $a_n = o(n^{\frac{1}{2}})$. *Then, as* $n \to \infty$, *with probability* 1, $l(\phi)$ *attains its minimum value at some point* $\tilde{\phi}$ *in the interior of the ball*

$\|\phi - \phi_0\| \le n^{-\frac{1}{3}}$ *and* $\tilde{\phi}$ *and* $\tilde{\lambda} = \lambda(\tilde{\phi})$ *satisfy* $Q_{1,n+1}(\tilde{\phi}, \tilde{\lambda}) = 0$ *and* $Q_{2,n+1}(\tilde{\phi}, \tilde{\lambda}) = 0$, *where* $Q_{1,n+1}(\phi, \lambda)$ *is defined in* (35) *and*

$$Q_{2n}(\phi, \lambda) = \frac{1}{n+1} \sum_{t=1}^{n+1} \frac{1}{1 + \lambda^\tau G_t(\phi)} \left(\frac{\partial G_t(\phi)}{\partial \phi} \right)^\tau \lambda. \tag{40}$$

The proof is similar to the proof of Lemma 1 of Qin and Lawless [28], so we omit the details.

Proof of Theorem 2.1 In what follows, we omit (ϕ, λ) in the notation if a function is evaluated at $(\phi_0, 0)$. Expanding $Q_{1,n+1}(\tilde{\phi}, \tilde{\lambda})$, $Q_{2,n+1}(\tilde{\phi}, \tilde{\lambda})$ at $(\phi_0, 0)$ leads to

$$0 = Q_{1,n+1}(\tilde{\phi}, \tilde{\lambda}) = Q_{1,n+1} + \left\{ \frac{\partial Q_{1,n+1}}{\partial \phi} \right\} (\tilde{\phi} - \phi_0) + \left\{ \frac{\partial Q_{1,n+1}}{\partial \lambda} \right\} \tilde{\lambda} + o_p(\delta_n) \tag{41}$$

and

$$0 = Q_{2,n+1}(\tilde{\phi}, \tilde{\lambda}) = Q_{2,n+1} + \left\{ \frac{\partial Q_{2,n+1}}{\partial \phi} \right\} (\tilde{\phi} - \phi_0) + \left\{ \frac{\partial Q_{2,n+1}}{\partial \lambda} \right\} \tilde{\lambda} + o_p(\delta_n), \tag{42}$$

where $\delta_n = \|\tilde{\phi} - \phi_0\|^2 + \|\tilde{\lambda}\|^2 = O_p(n^{-\frac{2}{3}})$.

Note that

$$\frac{\partial Q_{1,n+1}}{\partial \phi} = \frac{1}{n+1} \sum_{t=1}^{n+1} \frac{\partial G_t}{\partial \phi} = B + o_p(1), \tag{43}$$

$$\frac{\partial Q_{1,n+1}}{\partial \lambda} = -\frac{1}{n+1} \sum_{t=1}^{n+1} G_t G_t^\tau = -A + o_p(1), \tag{44}$$

$$\frac{\partial Q_{2,n+1}}{\partial \phi} = 0, \tag{45}$$

and

$$\frac{\partial Q_{1,n+1}}{\partial \lambda} = \frac{1}{n+1} \sum_{t=1}^{n+1} \frac{\partial G_t}{\partial \phi} = B^\tau + o_p(1). \tag{46}$$

These, combined with (41) and (42), give

$$\tilde{\lambda} = -\left\{ A^{-1} - A^{-1} B \left(B^\tau A^{-1} B \right)^{-1} B^\tau A^{-1} \right\} Q_{1,n+1} + o_p\left(n^{-\frac{1}{2}} \right) \tag{47}$$

and

$$\tilde{\phi} - \phi_0 = \left(B^\tau A^{-1} B \right)^{-1} B^\tau A^{-1} Q_{1,n+1} + o_p\left(n^{-\frac{1}{2}} \right). \tag{48}$$

Further, applying the central limit theorem to $Q_{1,n+1}$ and using Slustzky's theorem, we can prove Theorem 2.1. □

Proof of Theorem 2.2 Let $\tilde{\lambda}$ be the Lagrange multiplier corresponding to $\tilde{\phi}^{[s]}$, the maximum point of $l(\phi^{[s]})$. With this notation, we may write

$$l(s) = 2\sum_{t=1}^{n+1} \log\bigl\{1 + \tilde{\lambda}^\tau G_t^{[s]}\bigl(\tilde{\phi}^{[s]}\bigr)\bigr\}. \tag{49}$$

Note that

$$\tilde{\lambda}^\tau G_t^{[s]}\bigl(\tilde{\phi}^{[s]}\bigr) = \tilde{\lambda}^\tau G_t^{[s]} + \tilde{\lambda}^\tau \left\{\frac{\partial G_t^{[s]}}{\partial \phi^{[s]}}\right\}^\tau \bigl(\tilde{\phi}^{[s]} - \phi_0^{[s]}\bigr) + o_p(1). \tag{50}$$

This, together with (49), yields

$$\begin{aligned} ł(s) &= 2\tilde{\lambda}^\tau \sum_{t=1}^{n+1} G_t^{[s]} + 2\tilde{\lambda}^\tau \left\{\sum_{t=1}^{n+1} \frac{\partial G_t^{[s]}}{\partial \phi^{[s]}}\right\} \bigl(\tilde{\phi}^{[s]} - \phi_0^{[s]}\bigr) - \tilde{\lambda}^\tau \sum_{t=1}^{n+1} G_t^{[s]}\bigl(G_t^{[s]}\bigr)^\tau + o_p(1)\tilde{\lambda} \\ &= n^{-1} Q_{1,n+1}^\tau \bigl\{A^{-1} - A^{-1}B\bigl(B^\tau A^{-1}B\bigr)^{-1}B^\tau A^{-1}\bigr\} Q_{1,n+1} + o_p(1). \end{aligned}$$

Further note that $Q_{1,n+1}$ is asymptotic normal with covariance matrix A and $\{A^{-1} - A^{-1}B(B^\tau A^{-1}B)^{-1}B^\tau A^{-1}\}A\{A^{-1} - A^{-1}B(B^\tau A^{-1}B)^{-1}B^\tau A^{-1}\} = \{A^{-1} - A^{-1}B(B^\tau A^{-1}B)^{-1}B^\tau A^{-1}\}$. Therefore, we have $ł(s) \to \chi^2(p-k)$ in distribution as $n \to \infty$. The proof is complete. □

Proof of Theorem 2.3 Since $E(G_t(\phi)) \neq 0$, it follows that there exists $\delta > 0$, such that

$$\frac{1}{n}\sum_{t=1}^{n} G_t^\tau(\phi)\frac{1}{n}\sum_{t=1}^{n} G_t(\phi) - \delta^2 = o_p(1). \tag{51}$$

Furthermore, note that $E(G_t^\tau(\phi))^2 < \infty$. Thus, by a similar method to the proof of (27), we can prove that

$$\max_{1\le t\le n+1} \bigl\| G_t^\tau(\phi) \bigr\| = o_p\bigl(n^{\frac{1}{2}}\bigr). \tag{52}$$

Let $\check{\lambda} = n^{-\frac{2}{3}}(\frac{1}{n}\sum_{t=1}^{n} G_t(\phi)) \log n$. Then

$$\max_{1\le t\le n+1} \bigl|\check{\lambda}^\tau G_t(\phi)\bigr| = o_p(1). \tag{53}$$

Thus, with probability going to 1, $1 + \check{\lambda}^\tau G_t(\phi) > 0$ for $i = 1, \dots, n+1$. Using the duality of the maximization problem and (51)–(53), we have

$$\begin{aligned} l(\phi) &= \sup_{\lambda}\left(2\sum_{t=1}^{n+1} \log\bigl\{1 + \lambda^\tau G_t(\phi)\bigr\}\right) \ge 2\sum_{t=1}^{n+1} \log\bigl\{1 + \check{\lambda}^\tau G_t(\phi)\bigr\} \\ &= 2\sum_{t=1}^{n} \log\bigl\{1 + \check{\lambda}^\tau G_t(\phi)\bigr\} + o_p(1) = 2n^{\frac{1}{3}}\delta^2 \log(n) + o_p(1), \end{aligned}$$

which implies that $l(s) \to \infty$ in probability as $n \to \infty$. The proof is complete. □

Proof of Theorem 2.4. First, we consider EAIC. Consider the situation when s_0 is empty. Let $s = \{1\}$ which contains a single covariant. Based on expansion in the proof of Theorem 2.2, we can prove that $l(s_0) - l(s) \to \chi_1^2$, which implies that $\lim_{n\to\infty} P(l(s_0) - l(s) > 2) > 0$. Therefore, EAIC is not consistent.

Next, we consider EBIC. Suppose s is a model which does not contain s_0. Then, $E(G_t^{[s]}(\phi^{[s]})) \neq 0$ for any $\phi^{[s]}$. Therefore, we have $l(s) \geq 2n^{\frac{1}{3}}\delta^2 \log(n) + o_P(1)$. This order implies that

$$P\bigl(\mathrm{EBIC}(s) < \mathrm{EBIC}(s_0)\bigr) \leq P\bigl(l(s) - l(s_0) + p \log n\bigr) \to 0.$$

That is, EBIC will not select any model s that does not contain s_0.

Furthermore, if s contains s_0, and $k > 0$ additional insignificant variables, by Theorem 2.2, we have

$$l(s_0) - l(s) \to \chi_k^2,$$

which implies that

$$P\bigl(\mathrm{EBIC}(s) < \mathrm{EBIC}(s_0)\bigr) = P\bigl(l(s) - l(s_0) > k \log n\bigr) \to 0,$$

as $n \to \infty$. Thus, the model s will not be selected by EBIC as $n \to \infty$. Because p is finite, there are only finite number of $s$competing against s_0, and each of them has $o(1)$ probability being selection. So EBIC is consistent. The proof is complete. □

4 Conclusions

It should be pointed out that variable selection has always been an important problem for our statistician. Many variable selection methods have been proposed in the statistical literature. But for the variable selection method of GRCA, so far it has not been provided by statistician. In this paper, instead of parametric likelihood, we further propose an Akaike information criterion (EAIC) and a Bayesian information criterion (EBIC) for the variable selection problem of GRCA based on the empirical likelihood method. Moreover, we also prove that under some mild conditions the parametric EBIC is consistent, while the parametric EAIC is not when p is constant.

Acknowledgements
This work is supported by National Natural Science Foundation of China (No. 11571138, 11671054, 11301137, 11271155, 11371168, J1310022, 11501241), the National Social Science fund of China (16BTJ020), Science and Technology Research Program of Education Department in Jilin Province for the 12th Five-Year Plan (440020031139). "Thirteenth Five-Year Plan" Science and Technology Research Project of the Education of Jilin Province (Grant No. 2016103) and Jilin Province Natural Science Foundation (20130101066JC, 20130522102JH, 20150520053JH).

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All the authors read and approved the final manuscript.

Author details
[1]College of Mathematics, Jilin Normal University, Siping, P.R. China. [2]Public Foreign Languages Department, Jilin Normal University, Siping, P.R. China.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 27 May 2017 Accepted: 4 April 2018 Published online: 12 April 2018

References

1. Nicholls, D.F., Quinn, B.G.: Random Coefficient Autoregressive Models: An Introduction. Springer, New York (1982)
2. Tong, H.: Nonlinear Time Series. Oxford University Press, Oxford (1990)
3. Tong, H.: A note on a Markov bilinear stochastic process in discrete time. J. Time Ser. Anal. **2**, 279–284 (1981)
4. Feigin, P.D., Tweedie, R.L.: Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. J. Time Ser. Anal. **6**, 1–14 (1985)
5. Hwang, S.Y., Basawa, I.V.: Asymptotic optimal inference for a class of nonlinear time series models. Stoch. Process. Appl. **46**, 91–113 (1993)
6. Hwang, S.Y., Basawa, I.V.: Parameter estimation for generalized random coefficient autoregressive processes. J. Stat. Plan. Inference **68**, 323–327 (1998)
7. Hwang, S.Y., Basawa, I.V.: The local asymptotic normality of a class of generalized random coefficient autoregressive processes. Stat. Probab. Lett. **34**, 165–170 (1997)
8. Carrasco, M., Chen, X.: β-Mixing and moment properties of RCA models with application to GARCH(p,q). C. R. Acad. Sci., Sér. 1 Math. **331**, 85–90 (2000).
9. Zhao, Z.W., Wang, D.H.: Statistical inference for generalized random coefficient autoregressive model. Math. Comput. Model. **56**, 152–166 (2012)
10. Zhao, Z.W., Wang, D.H., Peng, C.X.: Coefficient constancy test in generalized random coefficient autoregressive model. Appl. Math. Comput. **219**, 10283–10292 (2013)
11. Theil, H.: Economic Forecasts and Policy. North-Holland, Amsterdam (1961)
12. Akaike, H.: A new look at the statistical model identification. IEEE Trans. Autom. Control **19**, 716–723 (1974)
13. Schwarz, G.: Estimating the dimension of a model. Ann. Stat. **6**, 461–464 (1978)
14. Mallows, C.L.: Some comments on C_p. Technometrics **15**, 661–675 (1973)
15. Hannan, E.J., Quinn, B.G.: The determination of the order of an autoregression. J. R. Stat. Soc., Ser. B, Stat. Methodol. **41**, 190–195 (1979)
16. Geweke, J., Meese, R.: Estimating regression models of finite but unknown order. Int. Econ. Rev. **16**, 55–70 (1981)
17. Cavanaugh, J.E.: A large-sample model selection criterion based on kull-backs symmetric divergence. Stat. Probab. Lett. **42**, 333–343 (1999)
18. Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.V.D.: Bayesian measures of model complexity and fit. J. R. Stat. Soc., Ser. B, Stat. Methodol. **64**, 583–639 (2002)
19. Tsay, R.S.: Order selection in nonstationary autoregressive models. Ann. Stat. **12**, 1151–1596 (1984)
20. Hurvich, C.M., Tsai, C.L.: Regression and time series model selection in small samples. Biometrika **76**, 297–307 (1989)
21. Pötscher, B.M.: Model selection under nonstationarity: autoregressive models and stochastic linear regression models. Ann. Stat. **17**, 1257–1274 (1989)
22. Andrews, D.W.K.: Consistent moment selection procedures for generalized method of moments estimation. Econometrica **67**, 543–564 (1999)
23. Andrews, D.W.K., Lu, B.: Consistent model and moment selection criteria for GMM estimation with applications to dynamic panel models. J. Econom. **101**, 123–164 (2001)
24. Hong, H., Preston, B., Shum, M.: Generalized empirical likelihood-based model selection criteria for moment condition models. Econ. Theory **19**, 923–943 (2003)
25. Zellner, A., Keuzenkamp, H.A., Mcaleer, M.: Simplicity, Inference and Modelling: Keeping It Sophisticatedly Simple. Cambridge University Press, Cambridge (2001)
26. Owen, A.B.: Empirical likelihood ratio confidence intervals for a single functional. Biometrika **75**, 237–249 (1988)
27. Owen, A.B.: Empirical Likelihood. Chapman and Hall, New York (2001)
28. Qin, J., Lawless, J.: Empirical likelihood and general estimating equations. Ann. Stat. **22**, 300–325 (1994)