Journal of Multivariate Analysis 124 (2014) 130–149

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Simultaneous confidence bands for sequential autoregressive fitting

Moritz Jirak

Humboldt-Universität zu Berlin, Institut für Mathematik, Unter den Linden 6, D-10099 Berlin, Germany

ARTICLE INFO

Article history:
Received 6 September 2012
Available online 5 November 2013

AMS 2000 subject classification:
60F05
62M10

Keywords:
Autoregressive process, asymptotic extreme value distribution
Yule–Walker estimators
Durbin–Levinson

ABSTRACT

Let $\{\mathbf{X}_k,\ k \in \mathbb{Z}\}$ be a zero mean causal AR(∞) process with parameter $\boldsymbol{\Theta} \in \mathbb{R}^\infty$. A very common fitting procedure is to employ the Yule–Walker equations in connection with the Durbin–Levinson algorithm, which yields the (recursive) sequence of estimators $\widehat{\boldsymbol{\Theta}}_m := (\widehat{\theta}_{m,1}, \ldots, \widehat{\theta}_{m,m})^\top$, $m = 1, 2, \ldots$. Under mild conditions, simultaneous confidence bands for $\widehat{\boldsymbol{\Theta}}_m, \widehat{\boldsymbol{\Theta}}_{m+1}, \ldots$ are derived. More precisely, it is shown that $\max_{d_n-\kappa_n \le m \le d_n} \max_{1\le h\le m} \left|\widehat{\theta}_{m,h} - \theta_h\right|$ converges to an extreme value distribution, where $d_n = \mathcal{O}\left(n^\delta\right)$, $\delta > 0$, and n denotes the sample size. The relation of κ_n and d_n depends on the bias term $\sum_{i=d_n-2\kappa_n}^{\infty} |\theta_i|$. This significantly extends a recent result in Jirak (2012). Moreover, extensions of results of An et al. (1982) and Bhansali (1978) are obtained. In addition, the behavior of Information criteria in the AR(∞) setting is briefly discussed.

1. Introduction

Let $\{X_k\}_{k\in\mathbb{Z}}$ be a stationary q-th order autoregressive process AR(q) with coefficient vector $\boldsymbol{\Theta}_q \in \mathbb{R}^q$, i.e., we have that

$$X_{k+1} = \sum_{i=1}^{q} \theta_i X_{k+1-i} + \epsilon_{k+1}, \tag{1.1}$$

where $\{\epsilon_k\}_{k\in\mathbb{Z}}$ is a zero mean IID-sequence with $\mathbb{E}(\epsilon_k^2) = \sigma^2$. When fitting such an AR(q) model, one usually chooses some order selection procedure for the order q, and an estimation method for the coefficients $\boldsymbol{\Theta}_q$. A considerable literature has evolved around these two issues, early pioneering contributions are due to Akaike [1,2], Mallows [33,34], Walker [45] and Yule [50], for more details see for instance [5,10,13,25,20,32]. Some of these approaches and related questions have then been further developed in [18,42,21,38,40,44], in particular the issue of order estimation has received considerable attention. Many of these methods are linked, in some way or other, to the confidence ellipsoids

$$\mathcal{M}_1(m) = \left\{\boldsymbol{\Theta}_m \in \mathbb{R}^m \,\middle|\, \left(\widehat{\boldsymbol{\Theta}}_m - \boldsymbol{\Theta}_m\right)^\top \widehat{\boldsymbol{\Gamma}}_m \left(\widehat{\boldsymbol{\Theta}}_m - \boldsymbol{\Theta}_m\right) \le n^{-1}\widehat{\sigma}_m^2 \chi^2_{1-\alpha}(m)\right\}, \tag{1.2}$$

where $\chi_{1-\alpha}{}^2(m)$ denotes the $(1-\alpha)$ quantile of the chi-square distribution with m degrees of freedom (cf. [5,13,20]), n denotes the number of observed realizations X_k, $1 \le k \le n$, and $\widehat{\boldsymbol{\Gamma}}_m, \widehat{\sigma}_m^2$ are estimators for the covariance matrix $\boldsymbol{\Gamma}_m$ and σ^2. One of the drawbacks of (1.2) is that it does not provide very useful inference for the single components $\left\{\left|\widehat{\theta}_i - \theta_i\right|\right\}_{1\le i\le d_n}$. Recently, in [30] a different confidence band was discussed,

$$\mathcal{M}_2(d_n) = \left\{\boldsymbol{\Theta}_{d_n} \in \mathbb{R}^{d_n} \,\middle|\, a_n^{-1}\left(n^{1/2} \max_{1\le i\le d_n} \left|(\widehat{\gamma}^*_{i,i})^{-1/2}(\widehat{\theta}_i - \theta_i)\right| - b_n\right) \le \sqrt{\widehat{\sigma}^2_{d_n} V_{1-\alpha}}\right\}, \tag{1.3}$$

E-mail address: m0ritz@yahoo.com.

0047-259X/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmva.2013.10.018

where $V_{1-\alpha}$ denotes the $1-\alpha$ quantile of a Gumbel-type distribution, and d_n increases with the sample size n (for the definition of a_n, b_n, d_n, $\widehat{\gamma}_{i,i}^*$ see Theorem 2.6). Clearly, $\mathcal{M}_2(d_n)$ provides more accessible (and tighter) bounds for the single elements $\{|\widehat{\theta}_i - \theta_i|\}_{1\le i\le d_n}$ when compared to $\mathcal{M}_1(d_n)$. This is for instance an important issue for so-called *subset* models, where one is interested in determining subsets $\mathcal{I}$ of significant parameters θ_i, $i \in \mathcal{I}$, and insignificant θ_j, $j \in \mathcal{J}$. For more details on this subject, see [37,12,35,36,23] and the references therein. Moreover, as pointed out in [30], $\mathcal{M}_{2(d_n)}$ is also very useful for order selection in sparse models.

However, the setup in [30] and also the confidence band $\mathcal{M}_2(d_n)$ itself suffer from two serious deficits. The first concerns the assumption of an underlying AR(d_n) process in [30]. As pointed out for instance by Burnham and Anderson [15], (see also [16]), it is more reasonable to assume that the data originates from an AR(∞) process. This is fueled by the fact that in many cases, an approximation with 'relatively' large lags provides superior results in prediction, see for instance [9,27,28, 26,43,46] and the references therein. Thereby, the assumption of an AR(d_n) process should be replaced by the assumption of an underlying AR(∞) process.

The second major concern regards the applicability of $\mathcal{M}_2(d_n)$ in practice, and was already pointed out in [30, Section 3]. Let us briefly elaborate on this particular issue. From an asymptotic point of view, it is enough to choose a large enough d_n, and then use $\mathcal{M}_2(d_n)$ to decide upon the redundance of single parameters θ_i and the order of the process $\{X_k\}_{k\in\mathbb{Z}}$. As stated in [30], this essentially boils down to 'just using the equation'

$$\widehat{\boldsymbol{\Theta}}_{d_n} = \widehat{\boldsymbol{\Gamma}}_{d_n}^{-1}\widehat{\boldsymbol{\Phi}}_{d_n}, \tag{1.4}$$

where $\widehat{\boldsymbol{\Phi}}_{d_n}$ denotes an estimate for the covariance vector $\boldsymbol{\Phi}_{d_n}$. However, the notion of *sufficiently large* d_n is difficult to handle in practice. Moreover, for large d_n, Eq. (1.4) may be numerically unstable due to the estimated inverse of the covariance matrix $\boldsymbol{\Gamma}_{d_n}$. This is a well-established fact in the literature (cf. [2,13,21,20,22]), and therefore usually all related estimators are computed in a *sequential manner* in practice (cf. [13,6,21,20,5]). For example, the successive values of Information criteria such as the AIC or BIC are normally based on the successive estimates $\widehat{\boldsymbol{\Theta}}_m = (\widehat{\theta}_{m,1}, \ldots, \widehat{\theta}_{m,m})^\top$, $m = 1, 2, \ldots$, and *not* the values $(\widehat{\theta}_{d_n,1}, \ldots, \widehat{\theta}_{d_n,m})^\top$, $m \le d_n$. Simply computing $\widehat{\boldsymbol{\Theta}}_{d_n}$ once for sufficiently large d_n and then using this particular vector to compute the AIC or BIC normally leads to significantly inferior results and cannot be advised. For the same reason, the usage of $\mathcal{M}_{3(d_n)}$ was advocated in [30, Section 3] in practice, which is the sequential analogue of $\mathcal{M}_2(d_n)$, and can be defined as

$$\mathcal{M}_3(d_n) = \left\{\boldsymbol{\Theta}_{d_n} \in \mathbb{R}^{d_n} \,\middle|\, a_n^{-1}\left(\sqrt{n}\max_{d_n-\kappa_n\le m\le d_n}\max_{1\le h\le m}(\widehat{\sigma}_m^2\widehat{\gamma}_{h,h}^*)^{-1/2}\,|\widehat{\theta}_{m,h}-\theta_h| - b_n\right) \le V_{1-\alpha}\right\},$$

where a_n, b_n and $V_{1-\alpha}$ are as in $\mathcal{M}_2(d_n)$. As our small simulation study in Section 4 shows, applying a sequential procedure relying on $\mathcal{M}_3(d_n)$ indeed gives substantially better results.

However, so far the theoretical justification for the actual usage of $\mathcal{M}_3(d_n)$ (i.e., whether or not the quantiles $V_{1-\alpha}$ actually fit) has not yet been established, and indeed this turns out to be far from trivial. Note that in $\mathcal{M}_2(d_n)$, the maximum is taken over d_n random variables. This is reflected in the normalizing sequences a_n, b_n, particularly in the centering sequence $b_n = \sqrt{2\log d_n}$, which mainly determines the growth rate of $\mathcal{M}_2(d_n)$. Contrary, in $\mathcal{M}_3(d_n)$, the number of involved random variables can be Cd_n^2, $C > 0$, hence one may expect a centering sequence $b_n = \sqrt{4\log d_n}$, thus an increase with factor $\sqrt{2}$. However, it turns out that the centering sequence b_n remains the same, i.e., $b_n = \sqrt{2\log d_n}$ (in fact $b_n = \sqrt{2\log\kappa_n}$). Clearly, this is not only a theoretical issue, since it tells us that the quantiles used for inference do not need to be amplified by the factor $\sqrt{2}$ in practice.

To the best of my knowledge, the only result in this direction was established by An et al. in [4], who proved a LIL-like behavior in case of ARMA(p, q) processes and $d_n = \mathcal{O}\big((\log n)^\delta\big)$, $\delta > 0$, see (2.6) for more details. The aim of this paper is to study the asymptotic behavior of $\mathcal{M}_3(d_n)$, thereby justifying its applicability. In doing so, we will also substantially extend the result of An et al.; stated in (2.6). Further, we also give an extension of a result of Bhansali [7] regarding the central limit theorem for $\sqrt{n}\big(\widehat{\boldsymbol{\Theta}}_{d_n} - \boldsymbol{\Theta}_{d_n}\big)$. In addition, the behavior of Information criteria in the AR(∞) setting is briefly discussed.

This paper is structured as follows. In Section 2 the main results are presented alongside some comments on the underlying assumptions. Further ramifications and comparisons to related results in the literature are given in Section 3. A simple simulation study to highlight the usage of $\mathcal{M}_3(d_n)$ is given in Section 4. Section 5 contains the proofs of the main results, the proofs of important technical lemmas are relegated to Section 6. Auxiliary results are provided in Section 7.

2. Main results

To state the results, we need to introduce some notation and specify the setting. We will always assume that we are given a sample X_k, $1 \le k \le n$ of size n, where $\{X_k\}_{k\in\mathbb{Z}}$ is a zero mean AR(∞) process given as in (1.1) with $q = \infty$. Note that in the literature one also encounters the notion $X_{k+1} + \psi_1 X_k + \cdots + \psi_j X_{k-j+1} + \cdots + = \epsilon_{k+1}$, which amounts to $\psi_i = -\theta_i$. Based on (1.1), we define the associated polynomial

$$\widetilde{\boldsymbol{\Theta}}(z) = 1 - \sum_{j=1}^{\infty}\theta_j z^j. \tag{2.1}$$