Journal of the Korean Statistical Society 46 (2017) 285–297

Contents lists available at ScienceDirect

Journal of the Korean Statistical Society

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

Tuning parameter selection for the adaptive LASSO in the autoregressive model

Sunghoon Kwon [a], Sangin Lee [b], Okyoung Na [c,*]

[a] Department of Applied Statistics, Konkuk University, Republic of Korea
[b] Department of Information and Statistics, Chungnam National University, Republic of Korea
[c] Department of Applied Information and Statistics, Kyonggi University, Republic of Korea

ARTICLE INFO

Article history:
Received 21 March 2016
Accepted 23 October 2016
Available online 17 November 2016

AMS 2000 Subject Classification:
62H99
62J07

Keywords:
Autoregressive model
Bayesian information criterion
Least absolute shrinkage and selection operator
Penalized estimation
Subset selection

ABSTRACT

We study the adaptive least absolute shrinkage and selection operator (LASSO) for the sparse autoregressive model (AR). Here, the sparsity of the AR model implies some of the autoregression coefficients are exactly zero, that must be excluded from the AR model. We propose the modified Bayesian information criterion (MBIC) as a way of selecting an optimal tuning parameter for the adaptive LASSO, which must be the most critical point in using the adaptive LASSO for the AR model. We prove that the adaptive LASSO obtained by minimizing the MBIC correctly distinguishes the true autoregression coefficients from zero asymptotically. The results hold even when the numbers of zero and nonzero true autoregression coefficients are diverging to infinity and the minimum of the absolute values of nonzero true autoregression coefficients decreases toward zero as the sample size increases. A small number of numerical studies are conducted to confirm the theoretical results.

1. Introduction

Consider the autoregressive (AR) model for the stochastic process $\{y_t\}$,

$$y_t = \sum_{j=1}^{p} \phi_j^* y_{t-j} + \varepsilon_t \tag{1}$$

for any integer t, where y_t is a target value of interest, ε_t is a random error, ϕ_j^* is a true autoregression coefficient, and p is a model order. A natural problem in using the AR model is the model identification when the AR model is sparse. Here, the sparsity of the AR model implies that there exists a non-empty subset $\mathcal{S}_T \subset \mathcal{S}_F = \{1, \ldots, p\}$ with $|\mathcal{S}_T| = q$ such that $\phi_j^* = 0$ for all $j \notin \mathcal{S}_T$, where $|\mathcal{A}|$ denotes the cardinality of a set $\mathcal{A}$. Hence the model identification becomes a procedure of recovering $\mathcal{S}_T$ and the true AR model,

$$y_t = \sum_{j \in \mathcal{S}_T} \phi_j^* y_{t-j} + \varepsilon_t, \tag{2}$$

from the full AR model (1) by using finite number of samples.

* Corresponding author.
E-mail address: okna@kyonggi.ac.kr (O. Na).

http://dx.doi.org/10.1016/j.jkss.2016.10.005
1226-3192/

There exist a number of literatures that study the model identification problem by using various information criteria. For example, Shibata (1976) proved that the final sub-model selected by the Akaike's information criterion (AIC) (Akaike, 1969) is asymptotically efficient under some mild conditions. However, the results also show that the final sub-model from the AIC is inconsistent in model identification but overestimates the unknown true model with a non-zero positive probability. Hannan and Quinn (1979) showed that the final sub-model obtained from the Bayesian information criterion (BIC) (Schwarz, 1978) and Hannan–Quinn criterion (HQC) (Hannan & Quinn, 1979) are consistent in model identification which is extended to the autoregressive moving average model by Hannan (1980). Tsay (1984) considered the AIC, BIC and HQC, simultaneously and generalized some asymptotic properties derived by Hannan (1980) and Shibata (1976) to the non-stationary AR model. Claeskens, Croux, and, Van Kerchhoven (2007) introduced the focussed information criterion (FIC) (Claeskens & Hjort, 2003) for the AR model as an alternative to the AIC to enhance prediction accuracy. A problem in using information criteria is how to construct candidate sub-models since the number of candidate sub-models increases exponentially fast as p does. For example, Sarkar and Kanjilal (1995) developed a way of using the AIC and BIC along to a sequence of sub-models produced by a singular value decomposition, which significantly reduces computational costs. Chen (1999) proposed to use a fully Bayesian framework to avoid overestimating the model which does not require any exhaustive search, and proved that the proposed method is consistent in model identification.

Another approach for the model identification problem in the AR model is the penalized estimation. The penalized estimation is one of the popular methods in the linear regression model (Fan & Li, 2001; Shao & Deng, 2012; Tibshirani, 1996; Zou, 2006) as well as generalized linear regression model (Fan & Peng, 2004; Kwon & Kim, 2012). The penalized estimation can determine the sparsity of the AR model and estimate corresponding non-zero coefficients simultaneously. Further, there exist many fast and efficient algorithms (Friedman, Hastie, Hofling, & Tibshirani, 2007; Kim, Choi, & Oh, 2008; Zou & Li, 2008) that can be applied to the AR model with large p. However, the most important property of the penalized estimation is that we need not to exhaustively search all the possible candidate sub-models, which significantly reduces computational costs. Further, the penalized estimators have nice asymptotic properties such as model identification consistency and minimax optimality (Fan & Peng, 2004; Kim et al., 2008; Leng, Lin, & Wahba, 2006; Raskutti, Wainwright, & Yu, 2011; Zhang, 2009; Zhang & Huang, 2008; Zhao & Yu, 2006; Zou, 2006). We refer to Zhang and Zhang (2012) for a well organized review of the penalized estimation for variable selection in the high-dimensional linear regression model.

There is a number of literatures that studies the penalized estimation for the AR model. For example, Nardi and Rinaldo (2011) studied the least absolute selection and shrinkage operator (LASSO) (Tibshirani, 1996) for the AR model and proved that the LASSO is consistent in parameter estimation and model identification under some regularity conditions. Schmidt and Makalica (2013) proposed the Bayesian LASSO by characterizing the model in terms of partial autocorrelations, and developed an efficient algorithm for computing the posterior mode by applying the coordinate descent algorithm. Sang and Sun (2013) studied two penalized estimations, the LASSO and smoothly clipped absolute deviation (SCAD) (Fan & Li, 2001), for the heavy-tailed AR model. They developed a one-step local linear approximation algorithm (Zou & Li, 2008) to implement the penalized maximum likelihood estimators, and proved that the final sub-model from the SCAD has the model identification consistency.

In this paper, we consider the adaptive LASSO (Zou, 2006) where the weight vector for the tuning parameter is available from the usual least squares estimation. The main contribution of the paper is developing a way of selecting tuning parameters, the modified Bayesian information criterion (MBIC), of which the minimizer recovers the true adaptive LASSO of the AR model asymptotically. Note that tuning parameter selection is the most important factor for the performance of penalized estimation. However, some usual ways of tuning parameter selection such as the K-fold cross validation and training-test random partition are not trivial or not theoretically supported for the AR model although they have been used in related literatures (Nardi & Rinaldo, 2011; Sang & Sun, 2013; Wang, Li, & Tsai, 2007a) without detailed discussions. Asymptotic properties of the adaptive LASSO and MBIC are given when the minimal absolute value of the true coefficient decreases, $\phi^*_{\min} = \min_{j\in\mathcal{S}_T} |\phi^*_j| \to 0$, and the model size increases, $p \to \infty$ and $p/n^2 \to 0$, as the sample size increases, $n \to \infty$. Hence, we can present the difficulty of model identification from the relationship among $\phi^*_{\min}$, p, q and n asymptotically. Various numerical studies confirm the theoretical results investigated in the paper.

The rest of the paper consists of the followings. Section 2 introduces the adaptive LASSO and the MBIC for the AR model. Sections 3 and 4 present asymptotic properties and Section 5 reports the simulation results. Discussions and technical details are provided in Section 6 and the Appendix, respectively.

2. Adaptive LASSO for the AR model

Given the samples y_t, $1-p \le t \le n$, from the AR model (1), let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$, where $\mathbf{X}_j = (y_{1-j}, \ldots, y_{n-j})^T$, $1 \le j \le p$, then we can rewrite the model as follows.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\phi}^* + \boldsymbol{\varepsilon},$$

where $\mathbf{y} = (y_1, \ldots, y_n)^T$, $\boldsymbol{\phi}^* = (\phi^*_1, \ldots, \phi^*_p)^T$ and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^T$. Given a weight vector $\mathbf{w} = (w_1, \ldots, w_p)^T$, we consider the adaptive LASSO proposed by Zou (2006):

$$\hat{\boldsymbol{\phi}}^{\lambda} = \arg\min_{\boldsymbol{\phi}} L_\lambda(\boldsymbol{\phi}), \tag{3}$$