Statistics and Probability Letters xx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Statistics and Probability Letters

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

Shrinkage estimation and variable selection in multiple regression models with random coefficient autoregressive errors

Saber Fallahpour [a,*], S. Ejaz Ahmed [b]

[a] Department of Mathematics and Statistics, University of Windsor, Windsor, ON, Canada
[b] Department of Mathematics, Brock University, St. Catharines, ON, Canada

ARTICLE INFO

Article history:
Received 5 December 2013
Received in revised form 25 May 2014
Accepted 28 May 2014
Available online xxxx

Keywords:
Penalty and shrinkage estimators
Multiple regression model
Random coefficient autoregressive error
Asymptotic bias and risk
Lasso
Adaptive lasso

ABSTRACT

In this paper, we consider improved estimation strategies for the parameter vector in multiple regression models with first-order random coefficient autoregressive errors (RCAR(1)). We propose a shrinkage estimation strategy and implement variable selection methods such as lasso and adaptive lasso strategies. The simulation results reveal that the shrinkage estimators perform better than both lasso and adaptive lasso when and only when there are many nuisance variables in the model.

1. Introduction

A classic problem in statistical analysis is finding a reasonable relationship between a response variable and a set of regressor variables under certain assumptions on the random errors. The usual assumption is that the errors are independent, identically distributed (i.i.d.) random variables. This has been later extended to many different cases when the errors are correlated. However, in many fields of research including economics, finance and biology it is well known that not all correlated error can be fitted well by linear time series errors. Therefore, much attention is now transferred to nonlinear time series models. Random coefficient time series models are one of the tools to handle the possible nonlinear features of real-life data. For a complete background on this model, we refer the reader to Nicholls and Quinn (1982). Liu and Tiao (1980) applied the random coefficient first-order autoregressive model to a panel data and they fitted this model to annual average hourly earnings in goods manufacturing in California. Also Singpurwalla and Soyer (1985) implemented this model in a real life data on software failures. They used this model for describing and assessing the software reliability growth or decay. In this paper, we consider an improved estimation for the parameters in a multiple regression model with random coefficient autoregressive errors. We consider methodologies for model selection and parameter estimation using shrinkage, lasso, and adaptive lasso strategies. Consider the following multiple regression model:

$$y_i = \mathbf{x}_i'\boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \ldots, n, \tag{1.1}$$

* Corresponding author.
E-mail address: fallahp@uwindsor.ca (S. Fallahpour).

http://dx.doi.org/10.1016/j.spl.2014.05.020
0167-7152/© 2014 Published by Elsevier B.V.

Please cite this article in press as: Fallahpour, S., Ahmed, S.E., Shrinkage estimation and variable selection in multiple regression models with random coefficient autoregressive errors. Statistics and Probability Letters (2014), http://dx.doi.org/10.1016/j.spl.2014.05.020

where y_i's are responses, $\mathbf{x}_i$ are known $p \times 1$ vectors of covariates, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$ is an unknown $p \times 1$ vector of regression parameters, and ε_i are unobservable random errors. Specifically we assume that ε_i is a first- order random coefficient autoregressive process, which is a stationary solution of

$$\boldsymbol{\varepsilon}_i = (\theta + z_i)\boldsymbol{\varepsilon}_{i-1} + e_i, \quad i = 1, \ldots, n, \tag{1.2}$$

where θ is the autoregression parameter and $\{z_i\}$ and $\{e_i\}$ are zero mean independent processes each consisting of i.i.d. random variables with finite second moments σ_z^2 and σ_e^2, respectively. Hwang and Basawa (1993, 1997) considered the estimation of the regression parameters of this model and established the local asymptotic normality for a class of generalized random coefficient autoregressive processes, respectively.

The main focus of this paper is to provide and compare different estimation strategies for model (1.1) with errors given in (1.2). We consider an adaptive shrinkage estimation strategy and suggest the shrinkage estimator (SE) and the positive shrinkage estimator (PSE). We study the properties of these estimators using the notion of Asymptotic Distribution Bias (ADB) and Asymptotic Distribution Risk (ADR). The shrinkage estimators are shown to have a higher efficiency than the classical estimators for a wide class of models (Fallahpour et al., 2012; Chitsaz and Ahmed, 2012). We also compare the relative performance of both lasso and adaptive lasso (AL) estimations with the SE and PSE. The performance of each estimator is evaluated in terms of simulated mean squared error. The simulation results show that the AL estimators have less mean squared error (MSE) than lasso. We also demonstrate that the shrinkage estimators outperform both lasso and AL estimators when, and only when, there are many nuisance variables in the model.

The rest of this paper is organized as follows. In Section 2, various estimation strategies, including a variable selection technique, are showcased. Section 3 provides asymptotic results of some estimators. In Section 4, we demonstrate via simulation that proposed strategy has good finite sample properties and are useful in practical applications. In Section 5, we provide a real data example. Finally, in Section 6, we present our concluding thoughts.

2. Variable selection and estimation strategies

Assuming model (1.1) with errors in (1.2), the generalized least squares (GLS) estimator of $\boldsymbol{\beta}$ is

$$\hat{\boldsymbol{\beta}}_G = (\mathbf{X}'\boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{y},$$

where

$$\boldsymbol{\Gamma}^{-1}(\theta) = \begin{pmatrix} 1 & -\theta & 0 & 0 & \cdots & 0 \\ -\theta & 1+\theta^2 & -\theta & & & \vdots \\ 0 & -\theta & 1+\theta^2 & -\theta & & \\ \vdots & & & & & \\ 0 & & \cdots & & -\theta & 1 \end{pmatrix}$$

and θ is unknown, as is often the case in practice; then $\boldsymbol{\Gamma}(\theta)$ is replaced by $\boldsymbol{\Gamma}(\hat{\theta})$ where $\hat{\theta}$ is the least squares estimator of θ based on the residuals $\hat{\varepsilon}_i = y_i - \mathbf{x}_i'\hat{\boldsymbol{\beta}},\ i = 1, \ldots, n$, and is given by $\hat{\theta} = \sum_{i=2}^{n} \hat{\varepsilon}_i\hat{\varepsilon}_{i-1} / \sum_{i=2}^{n} \hat{\varepsilon}_{i-1}^2$. Note that $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ is the ordinary least square (OLS) estimate of $\boldsymbol{\beta}$. The properties of this estimator were investigated in Hwang and Basawa (1993).

Let us consider the following partition of $\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$, where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ have dimensions p_1 and p_2 respectively, and $p_1 + p_2 = p,\ p_i \geq 0$ for $i = 1, 2$. Thus we can rewrite the model (1.1) as follows:

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\varepsilon}.$$

The unrestricted generalized least square estimator (UE) $\hat{\boldsymbol{\beta}}_1$ of $\boldsymbol{\beta}_1$ will be in the form of

$$\hat{\boldsymbol{\beta}}_{1G} = (\mathbf{X}_1' M_{\boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{X}_2}\mathbf{X}_1)^{-1}\mathbf{X}_1' M_{\boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{X}_2} y,$$

where $\mathbf{X}_1$ is composed of the first p_1 column vectors of $\mathbf{X}$, $\mathbf{X}_2$ is composed of the last p_2 column vectors of $\mathbf{X}$ and

$$M_{\boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{X}_2} = \boldsymbol{\Gamma}^{-1}(\hat{\theta}) - \boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{X}_2(\mathbf{X}_2'\boldsymbol{\Gamma}^{-1}(\hat{\theta})\mathbf{X}_2)^{-1}\mathbf{X}_2'\boldsymbol{\Gamma}^{-1}(\hat{\theta}).$$

We are mainly interested in the estimation of $\boldsymbol{\beta}_1$ when based on some variable selection method or prior information from previous studies it is plausible that $\boldsymbol{\beta}_2$ is close to some specified $\boldsymbol{\beta}_2^0$ which, without loss of generality, we may set it to $\mathbf{0}$ and our goal in this paper is to construct an efficient estimation for the regression parameter $\boldsymbol{\beta}_1$ when $\boldsymbol{\beta}_2 = \mathbf{0}$. For example, in the case of a multifactor design, we may be interested in estimating the main effects $\boldsymbol{\beta}_1$, while there is a question whether Q3 the vector of interaction effects $\boldsymbol{\beta}_2$ may be ignored. Now suppose $\boldsymbol{\beta}_1$ is the $p_1 \times 1$ coefficient vector for main effects and $\boldsymbol{\beta}_2$ is the $p_2 \times 1$ coefficient vector for nuisance effects and there is evidence that nuisance variables do not provide useful information, that is, $\boldsymbol{\beta}_2 = 0$. By removing these variables, we have a candidate submodel as

$$\mathbf{y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon}. \tag{2.1}$$

Please cite this article in press as: Fallahpour, S., Ahmed, S.E., Shrinkage estimation and variable selection in multiple regression models with random coefficient autoregressive errors. Statistics and Probability Letters (2014), http://dx.doi.org/10.1016/j.spl.2014.05.020