Economics Letters 148 (2016) 68–71

Contents lists available at ScienceDirect

Economics Letters

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

Model averaging with high-dimensional dependent data

Shangwei Zhao [a], Jianhong Zhou [b,*], Hongjun Li [c]

[a] College of Science, Minzu University of China, China
[b] Department of Credit Management, Guangdong University of Finance, China
[c] ISEM, Capital University of Economics and Business, China

ARTICLE INFO

Article history:
Received 7 July 2016
Received in revised form 13 September 2016
Accepted 14 September 2016
Available online 19 September 2016

JEL classification:
C2
C13

Keywords:
Dependent data
High dimension
Model averaging
Optimality

ABSTRACT

The past two decades witnessed a prosperous literature on model averaging, however, few authors have examined model averaging under high-dimensional data setting. An exception is Ando and Li (2014), which proposed a model averaging procedure to improve prediction accuracy under high-dimensional independent data setting. In this paper, we broaden Ando and Li's scope of analysis to allow dependent data. We show that under the dependent data setting, their model averaging estimator is still asymptotically optimal. Simulation study demonstrates the finite sample performance of the estimator in a variety of dependent data settings.

1. Introduction

In the past two decades, model averaging has been well developed. Contributions to model averaging come from different data settings, such as independent and homoskedastic setting (Hansen, 2007; Zhang et al., 2015; Zhao, 2014), independent and heteroskedastic setting (Hansen and Racine, 2012; Liu and Okui, 2013; Magnus et al., 2011; Zhao et al., 2016), dependent setting (Gao et al., 2016; Zhang et al., 2013), censoring data (Hjort and Claeskens, 2006; Zhang et al., 2012), and missing data (Schomaker et al., 2010; Zhang, 2013). However, in high dimension situation, almost all methods developed in these papers will be infeasible from computational perspective because theoretically there are 2^p candidate models where p is the number of regressors. When p is 20, 2^p will be above one million.

Recently, Ando and Li (2014) developed a model averaging procedure using Jackknife model averaging (JMA) proposed by Hansen and Racine (2012). The marginal correlation between each predictor and the response variable was used to partition predictors into several groups and then they used these groups to prepare candidate models and chose the weight vector by minimizing a cross-validation criterion. Without requiring the weights sum to one, they proved the asymptotic optimality of their procedure. Simulation results show that the procedure performs much better than the existing model selection and averaging methods. But in Ando and Li (2014), the data observations are requested to be independent.

In the current paper, we further investigate the procedure of Ando and Li (2014) under dependent data setting. We show that under this setting, their model averaging estimator is still asymptotically optimal and has promising finite sample performances.

2. Model and estimation

We follow Ando and Li (2014) notations as much as possible for readers' convenience and consider the multiple linear regression model

$$y = \sum_{j=1}^{p} \beta_j x_j + \epsilon, \tag{1}$$

where y is the response variable, $x_1, \ldots, x_p$ are explanatory variables, p is allowed to increase with the sample size n and even larger than n, and the random error term ϵ has mean $E[\epsilon] = 0$ and $var[\epsilon] = \Sigma$. In Ando and Li (2014), Σ is only allowed to be a diagonal matrix, but here we allow Σ to be a general non-negative definite matrix, i.e., we consider both potentially heteroscedasticity and serial correlation in the random error term.

Following Ando and Li (2014), we prepare candidate models using the marginal correlation between each explanatory variable

* Corresponding author.
E-mail address: jianhzhou86@163.com (J. Zhou).

http://dx.doi.org/10.1016/j.econlet.2016.09.010

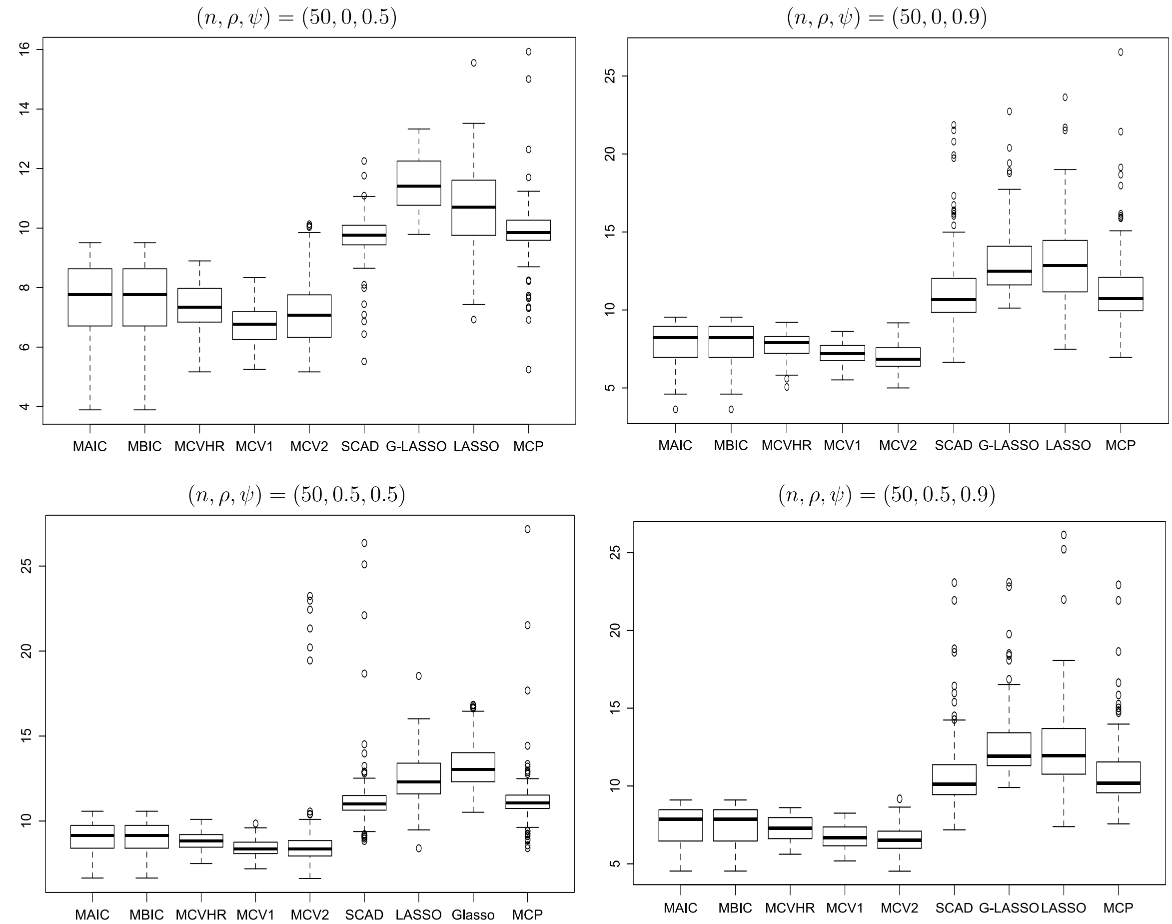

Fig. 1. Simulation results: MSE with $n = 50$.

and the response variable. We partition the p marginal correlations into $M + 1$ groups by the absolute values of the marginal correlation. The first group has the highest values and the $M+1$ group has values closest to zero. We drop the $M + 1$ group. Thus the number of models is M.

Let model M_k consist of the regressors with marginal correlations falling into the kth group. Denote the kth candidate model as

$$y = \sum_{j \in A_k} \beta_j x_j + \epsilon, \qquad k = 1, \ldots, M \tag{2}$$

where A_k is the index set of regressors to be included in model M_k. Denote X_k as the regressor matrix and β_k as the p_k-dimensional parameter vector under model M_k. Let $\mu = E(y)$. We assume $p_k \leq n$ and X_k has full column rank. Using least square estimation, $\widehat{\beta}_k = (X_k' X_k)^{-1} X_k' y$ and the least squares prediction $\widehat{\mu}_k = X_k \widehat{\beta}_k$. Denote the projection matrix $X_k (X_k' X_k)^{-1} X_k'$ by H_k. Let the M-dimensional weight vector $w = (w_1, \ldots, w_M)'$ come from $Q_n = \left\{ w \in [0, 1]^M : 0 \leq w_k \leq 1 \right\}$. The model average predictor μ can be written as

$$\begin{aligned} \widehat{\mu}(w) &= \sum_{k=1}^{M} w_k \widehat{\mu}_k = \sum_{k=1}^{M} w_k X_k (X_k' X_k)^{-1} X_k' y \\ &= \sum_{k=1}^{M} w_k H_k y = H(w) y, \end{aligned} \tag{3}$$

where $H(w) = \sum_{k=1}^{M} w_k H_k$. Note that the restriction $\sum_{k=1}^{M} w_k = 1$ is unnecessary in this paper.

We estimate the weights using the delete-one cross-validation approach (i.e., JMA) as used in Hansen and Racine (2012). Let $\widetilde{\mu}_k^{(-\alpha)}$ be the predicted value of the αth observation from model M_k and $\widetilde{\mu}_k = (\widetilde{\mu}_k^{(-1)}, \ldots, \widetilde{\mu}_k^{(-n)})'$. Then, we can write $\widetilde{\mu}_k = \widetilde{H}_k y$, where $\widetilde{H}_k = D_k (H_k - I) + I$, $D_k = \operatorname{diag}\left\{ (1 - h_{k1})^{-1}, \ldots, (1 - h_{kn})^{-1} \right\}$, and $h_{k\alpha}$ is the αth diagonal element of H_k. Then the delete-one model averaging predictor of μ is

$$\widetilde{\mu}(w) = \sum_{k=1}^{M} w_k \widetilde{\mu}_k = \sum_{k=1}^{M} w_k \widetilde{H}_k y = \widetilde{H}(w) y, \tag{4}$$

where $\widetilde{H}(w) = \sum_{k=1}^{M} w_k \widetilde{H}_k$. The cross-validation criterion is $CV(w) = \| y - \widetilde{\mu}(w) \|^2$. We select the weight vector w as

$$\widehat{w} = \arg\min_{w \in Q_n} CV(w). \tag{5}$$

Next, we show the asymptotic optimality of $\widehat{w}$ with a possibly non-diagonal Σ. Let $L(w) = \| \widehat{\mu}(w) - \mu \|^2$, $R(w) = E\{L(w)\}$, $\zeta_n = \inf_{w \in Q_n} R(w)$, and $\bar{\lambda}(\Upsilon)$ and $\lambda(\Upsilon)$ denote the maximal diagonal element and the maximal singular value of matrix Υ, respectively. All limiting processes discussed in this and subsequent sections are as $n \to \infty$. We need the following regularity conditions.

Condition (C.1). $\max_{k=1,\ldots,M} p_k^{-1} \bar{\lambda}(H_k) = O(n^{-1})$.

Condition (C.2). $\|\mu\|^2 = O(n)$.

Condition (C.3). $\lambda(\Sigma) = O(1)$.

Condition (C.4). $M^3 \zeta_n^{-2} n = o(1)$ *and* $M^2 \zeta_n^{-1} \max_{k=1,\ldots,M} p_k = o(1)$.