Computational Statistics and Data Analysis 96 (2016) 104–119

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Regularized estimation for the least absolute relative error models with a diverging number of covariates

Xiaochao Xia [a,∗], Zhi Liu [b,c], Hu Yang [a]

[a] College of Mathematics and Statistics, Chongqing University, China
[b] Department of Mathematics, University of Macau, Macau
[c] UMacau Zhuhai Research Institute, China

ARTICLE INFO

Article history:
Received 1 October 2014
Received in revised form 27 October 2015
Accepted 29 October 2015
Available online 5 November 2015

Keywords:
Variable selection
Diverging number of covariates
Least absolute relative error
Least squares approximation
Oracle properties

ABSTRACT

This paper considers the variable selection for the least absolute relative error (LARE) model, where the dimension of model, p_n, is allowed to increase with the sample size n. Under some mild regular conditions, we establish the oracle properties, including the consistency of model selection and the asymptotic normality for the estimator of non-zero parameter. An adaptive weighting scheme is considered in the regularization, which admits the adaptive Lasso, SCAD and MCP penalties by linear approximation. The theoretical results allow the dimension diverging at the rate $p_n = o(n^{1/2})$ for the consistency and $p_n = o(n^{1/3})$ for the asymptotic normality. Furthermore, a practical variable selection procedure based on least squares approximation (LSA) is studied and its oracle property is also provided. Numerical studies are carried out to evaluate the performance of the proposed approaches.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

In this paper, we consider the following multiplicative model:

$$Y = \exp(X^T \boldsymbol{\beta})\varepsilon, \tag{1}$$

where Y is the response variable, X is a p-dimensional vector of covariates, $\boldsymbol{\beta}$ is an unknown p-vector of parameters, and ε is random noise. Both Y and ε considered in model (1) are positive variables. When the response variable Y is a failure time, model (1) is referred to the accelerated failure time (AFT) model, which has been widely applied in survival analysis. For instance, under the AFT model, estimation of $\boldsymbol{\beta}$ has been documented in Jin et al. (2003), Huang et al. (2006), Huang and Ma (2010), among others.

A mostly recent work regarding estimation of parameter $\boldsymbol{\beta}$ of model (1) is due to Chen et al. (2010), who proposed to estimate the parameter $\boldsymbol{\beta}$ via minimizing the following objective function

$$L_n(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left\{ \left| \frac{y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{y_i} \right| + \left| \frac{y_i - \exp(\mathbf{x}_i^T \boldsymbol{\beta})}{\exp(\mathbf{x}_i^T \boldsymbol{\beta})} \right| \right\}, \tag{2}$$

where $(y_i, \mathbf{x}_i)_{i=1}^n$ are independent observations of (Y, X) of model (1). A sum of the absolute relative errors is considered in the empirical objective function (2). Such criterion is primarily motivated by following considerations. On the one hand, it is

∗ Corresponding author.
E-mail address: xxc@cqu.edu.cn (X. Xia).

http://dx.doi.org/10.1016/j.csda.2015.10.012
0167-9473/© 2015 Elsevier B.V. All rights reserved.

more attractive to consider the relative errors than the absolute errors in many practical applications. To name a few, Narula and Wellington (1977) proposed an estimate by using the absolute relative errors in linear model. Makridakis et al. (1984) studied the relative errors in the analysis of time series data. Khoshgoftaar et al. (1992) established the strong consistency of the estimators in the case of squared relative errors and absolute relative errors, respectively, in nonlinear regression model. Park and Stefanski (1998) considered the best mean squared relative error prediction. On the other hand, only considering the ratio of the error relative to some target value tends to be more or less restrictive. For instance, assume that y_i takes a large value (say 50) and $\exp(\mathbf{x}_i^T\boldsymbol{\beta})$ takes small (say 5), then when one chooses y_i as the target, the relative error is $\left|\frac{y_i-\exp(\mathbf{x}_i^T\boldsymbol{\beta})}{y_i}\right| = 0.9$, whereas, the value is magnified ten times when $\exp(\mathbf{x}_i^T\boldsymbol{\beta})$ is viewed as the target. This causes that using either of these two relative errors as the final loss function in the model is inadequate. In particular, the stock buyers could regard the market price as the target value, while the stock sellers could regard the intrinsic value of a stock as the target. Hence, a combination of these two kinds of relative errors may be preferred in certain circumstances like this stock price data. More discussions on the choice of relative errors are referred to Ye (2007) for accounting modeling. A natural and reasonable tradeoff is to consider the sum of these two relative errors, which is called the least absolute relative errors (LARE) loss, due to Chen et al. (2010). One advantage of making use of such loss function is that it is scale free or unit free. This might be particularly useful in applying the LARE criterion to analyze the data with certain types. To see the problem of minimizing $L_n(\beta)$ in (2) more precisely, if the response comes from a log normal distribution, the minimizer behaves like the (conditional) sample median for a large sample size (or asymptotically), whereas if the response is from a distribution with efficient density (see Section 5.1), it is just the MLE. There are some recent works documented about the inference of model (1) with the LARE loss in the literature. Besides Chen et al. (2010), Li et al. (2014) employed the empirical likelihood technique to make the statistical inference and built a nonparametric version of Wilks' theorem. Zhang and Wang (2013) extended model (1) to a partially linear multiplicative model and investigated the asymptotic results for the parametric and nonparametric estimates using a locally weighted LARE loss function.

Instead of considering the estimation, the purpose of this paper is to study the problem of variable selection raised by model (1). Over the past two decades, model selection has been becoming a very important problem and extensively addressed in modern statistics. A large body of relevant literatures regarding various regularization methods were developed for both the parametric and nonparametric models, see for example, Lasso by Tibshirani (1996), SCAD by Fan and Li (2001), adaptive Lasso (ALasso) by Zou (2006), sparsity–smoothness penalty by Meier et al. (2009) in generalized additive models, adaptive group Lasso by Huang et al. (2010) for nonparametric additive models, MCP by Zhang (2010) and robust adaptive method by Fan et al. (2014a). To the best of our knowledge, most of the efforts contributed to variable selection concentrate on the commonly encountered loss functions, such as square error loss (ℓ_2) for LS (Fan and Li, 2001; Zou, 2006; Wang and Leng, 2007), absolute error loss (ℓ_1) for LAD (Wang et al., 2007), quantile loss for quantile regression (Wang et al., 2012; Fan et al., 2014a), logistic loss in a binary classification model (Breheny and Huang, 2015). However, the variable selection for model (1) with the LARE loss in the sense of diverging number of covariates has not been considered yet. Hence, we would fill the gap by this paper.

The contributions of this paper can be summarized as follows. First, we develop a variable selection procedure to select and estimate the important parameters in model (1) simultaneously. The number of parameters considered in model (1) is allowed to grow with the sample size. This kicks off a moderate-high dimensional multiplicative data analysis, and generalizes the results of Zhang and Wang (2013), where they assume the dimension is fixed. Second, we incorporate the LARE loss into the empirical objective function, which is significantly distinct from other loss functions as stated previously. Third, a general adaptive weighting scheme is studied. It allows for some folded concave penalty functions via linear approximation. Fourth, an efficient variable selection procedure based on the least squares approximation is provided for practical application. This generalizes the result of Wang and Leng (2007) in the sense of high dimension. The simulation studies and real data analysis provide some empirical evidence for the correctness of the proposed approaches.

The rest of paper is organized in the following way. In Section 2, we introduce the estimation methodology and propose a regularized estimator. In Section 3, some regular conditions are given and some asymptotic results of the estimator are presented. In Section 4, we propose a practical variable selection procedure. The simulation studies and a real data analysis are conducted in Section 5. We conclude the paper in Section 6. All technical proofs are relegated to the Appendix.

2. Methodology

For the ease of statement, we make some notations first. Denote by $\mathbf{y} = (y_1, \ldots, y_n)^T$ an n-dimensional observation vector of response variable. Let $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T = (\tilde{\mathbf{x}}_1, \ldots, \tilde{\mathbf{x}}_p)$ be an $n \times p$ design matrix and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)^T$ the parameter vector. Throughout this paper, assume that $\mathbf{X}$ is a fixed design matrix. To enhance the accuracy of model fitting and prediction as well as the interpretability, we make an assumption that only a small proportion of covariates contribute to the response variable, that is sparsity, which is a common faith in the context of model selection, see Fan and Li (2001) and Zou (2006). Suppose that the number of nonzero parameters is s. From now on, we use the subscript n to emphasize the dependence between the dimension p and the sample size n. Thus, p_n and s_n are used to stand for the dimension and sparsity of the model, respectively, whenever the context is clear.

Without loss of generality, we assume the true parameter of model (1) is $\boldsymbol{\beta}^* = (\boldsymbol{\beta}_1^{*T}, \mathbf{0}^T)^T$ with $\boldsymbol{\beta}_1^* \in \mathbf{R}^{s_n}$. This means that only the first s_n parameters are non-vanishing. Denote by the true model $\mathcal{M}_* = \text{supp}(\boldsymbol{\beta}^*) = \{1, \ldots, s_n\}$ and $\mathcal{M}_*^c$ its

Download English Version:

https://daneshyari.com/en/article/416388

Download Persian Version:

https://daneshyari.com/article/416388

Daneshyari.com