Computational Statistics and Data Analysis 109 (2017) 105–120

Contents lists available at ScienceDirect

Computational Statistics and Data Analysis

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

Inference for biased transformation models

Xuehu Zhu [a,b], Tao Wang [c], Junlong Zhao [d], Lixing Zhu [d,e,*]

[a] School of Mathematics and Statistics, Xi'an Jiaotong University, Xi'an, China
[b] Qilu Securities Institute for Financial Studies, Shandong University, China
[c] Department of Bioinformatics and Biostatistics, Shanghai Jiaotong University, Shanghai, China
[d] School of Statistics, Beijing Normal University, Beijing, China
[e] Department of Mathematics, Hong Kong Baptist University, Hong Kong

ARTICLE INFO

Article history:
Received 14 April 2016
Received in revised form 10 November 2016
Accepted 10 November 2016
Available online 23 November 2016

Keywords:
Estimation consistency
Linear transformation models
Model bias correction
Non-sparse structure

ABSTRACT

Working regression models are often parsimonious for practical use and however may be biased. This is because either some strong signals to the response are not included in working models or too many weak signals are excluded in the modeling stage, which make cumulative bias. Thus, estimating consistently the parameters of interest in biased working models is then a challenge. This paper investigates the estimation problem for linear transformation models with three aims. First, to identify strong signals in the original full models, a sufficient dimension reduction approach is applied to transferring linear transformation models to pro forma linear models. This method can efficiently avoid high-dimensional nonparametric estimation for the unknown model transformation. Second, after identifying strong signals, a semiparametric re-modeling with some artificially constructed predictors is performed to correct model bias in working models. The construction procedure is introduced and a ridge ratio estimation is proposed to determine the number of these predictors. Third, root-n consistent estimators of the parameters in working models are defined and the asymptotic normality is proved. The performance of the new method is illustrated through simulation studies and a real data analysis.

1. Introduction

Consider the linear transformation model:

$$H(Y) = X^{\top}\beta + \epsilon, \tag{1}$$

where Y is the response variable, $H(\cdot)$ denotes an unknown monotone transformation function, $X = (X_1, X_2, \ldots, X_p)^{\top}$ is the p-dimensional predictor vector having mean zero, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)^{\top}$ is the regression parameter vector of interest, and the error ϵ, which is independent of X, has a continuous distribution. As H is unknown, without loss of generality, we assume that the L_2 norm of β is one. This model, which circumvents the so-called curse of dimensionality, is widely used in statistical analysis.

In practical use, it is often not to contain all the predictors in a working model as when there are many predictors, we often remove those 'weak' signals for the response such that a parsimonious working model can be built. This is particularly the

* Corresponding author at: School of Statistics, Beijing Normal University, Beijing, China.
E-mail address: lzhu@hkbu.edu.hk (L. Zhu).

http://dx.doi.org/10.1016/j.csda.2016.11.008
0167-9473/

case in high-dimensional paradigms because to well estimate the parameter of interest β, variable selection is indispensable and hopefully only those 'strong' signals are retained in the working model. Here 'strong' signals are reflected by the magnitudes of the corresponding coefficients β_i. To have estimation consistency, model sparsity is an usual assumption. That is, a small portion of the components of β are assumed to be significantly non-zero and the others are zero. However, this is often hard to judge. In practice, it is often that many weak signals with small coefficients, but non-zero coefficients in the model are removed out from the regression function. When the number of these predictors is large, the cumulative bias of working model is often not negligible. Further the selection process would also unwittingly rule out some strong signals. All these cause the difficulty of having estimation consistency. Relevant references includes Leeb and Pötscher (2005) and Zhang and Huang (2008) who have observed similar problems. For semiparametric or nonparametric modeling, parsimonious modeling is even more important, otherwise, consistent estimation is harder to achieve because nonparametric estimation would be involved. The linear transformation model is an example as the function $H(\cdot)$ is nonparametric.

Lin et al. (2016) investigated this issue for linear models. They proposed a bias correction method to achieve estimation consistency via a semiparametric remodeling. Their method highly relies on the linear model structure, and the two methods proposed by Lin et al. (2016), which select quasi-instrumental variables for the semiparametric remodeling, would be complex to implement and theoretically do not have the consistency of the estimated number of quasi-instrumental variables. For the linear transformation model under study, the problem is more difficult. First, as the model transformation function H is unknown and when X is high-dimensional, it is not a wise idea to use a nonparametric method to estimate the transformation function H and then the parameter of interest. To have an estimation procedure without estimating the transformation function H, our idea used in this paper is to apply a sufficient dimension reduction approach to transferring the original model to an equivalent pro forma 'linear' model. With this new formulation, estimating the transformation function H is completely averted. The key for this method to work is the equivalence of strong and weak signals between the original nonparametric transformation and the pro forma linear model. This will be confirmed in the next section. Therefore, for bias correction, we will work on the pro forma linear model and then, if necessary, establish a new transformation (may be still nonparametric) to fit data. It is worth mentioning that unlike the classical linear model, given the selected predictors, the conditional expectation of the error in this pro forma 'linear' model is not centered, while it just has a zero expectation of the product of the error and the predictors. An idea in spirit similarly as that of Lin et al. (2016) is used to construct some 'new' predictors such that a semiparametric re-modeling is possible to remove bias. However, as the linearity is just pro forma, the bias correction is more delicate than that in Lin et al. (2016). The detail will be elaborated in Section 3. Second, we suggest a ridge ratio estimation (RR) to estimate the number of those constructed predictors. The consistency will be proved.

We now formulate the model. Assume for simplicity that the first q predictors, represented by $\mathcal{I} = \{1, 2, \ldots, q\}$, are relatively strong signals, and the last $p - q$ predictors are weak signals but not necessarily irrelevant to the response. We rewrite X and β as $X = (Z^\top, U^\top)^\top$ and $\beta = (\beta_{\mathcal{I}}^\top, \beta_{\mathcal{I}^c}^\top)^\top$, respectively, where $Z = (X_1, \ldots, X_q)^\top$, $U = (X_{q+1}, \ldots, X_p)^\top$ and $\mathcal{I}^c$ is the compliment of $\mathcal{I}$ in the complete set $\{1, \ldots, p\}$. Correspondingly, the full model is rewritten as

$$H(Y) = Z^\top \beta_{\mathcal{I}} + \tilde{\epsilon}, \tag{2}$$

where $\tilde{\epsilon} = U^\top \beta_{\mathcal{I}^c} + \epsilon$ denotes the error term in this working model. A typical example is a working model that is based on a variable selection procedure such as the LASSO (Tibshirani, 1996). As commented above, when some strong signals are not included in the regression part or when the full model (1) is not sparse, the working model (2) could be biased:

$$E(\tilde{\epsilon}|Z) = \sum_{j \notin \mathcal{I}} \beta_j E(X_j|Z) \neq 0. \tag{3}$$

Consequently, it is of importance to obtain a consistent estimate of $\beta_{\mathcal{I}}$ or more precisely the normalized $\beta_{\mathcal{I}}$ as in the nonparametric transformation model, the scale of $\beta_{\mathcal{I}}$ is not identifiable. Without notational confusion, we still use $\beta_{\mathcal{I}}$ as its normalized one. The goal of the present paper is to construct a bias-corrected model and to obtain a root-n consistent estimate of $\beta_{\mathcal{I}}$. Our idea is to first transfer the linear transformation model to an equivalent pro forma 'linear' model such that the modeling becomes easier without involving nonparametric estimation for the model transformation H. After that, we introduce a new method to conveniently determine the number of "new" predictors for the purpose of bias correction. The semiparametric re-modeling can then be implemented for the purpose of estimation consistency.

The rest of this paper is organized as follows. In Section 2, we transfer the original model as an equivalent pro forma linear model. We then give the definition of the index set of relatively strong signals in a non-sparse transformation model and the identifiability of the index set. In Section 3, after introducing a vector of 'new' predictors, a partially linear remodeling is implemented in which the parameters of interest are in the linear component when both the response and regression function are re-centered. The ridge ratio method to determine the number of 'new' predictors is also suggested in this section and the asymptotic properties are derived. It is noteworthy that the identifiability of the index set is mainly for theoretical purpose such that the selected model is non-random. For practical use, it is not necessary. This can be seen in Section 3. Numerical simulations are conducted to examine the performance of the new method in Section 4. All the proofs are postponed to Appendix.