Statistical Methodology 25 (2015) 51–58

Contents lists available at ScienceDirect

Statistical Methodology

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

Adaptive testing for the partially linear single-index model with error-prone linear covariates

Zhensheng Huang [a,b,*], Quanxi Shao [b], Zhen Pang [c], Bingqing Lin [d]

[a] *School of Science, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, PR China*
[b] *CSIRO Digital Productivity Flagship, Private Bag 5, Wembley, WA 6913, Australia*
[c] *Department of Applied Mathematics, the Hong Kong Polytechnic University, Hong Kong*
[d] *Institute of Statistical Science, Shenzhen University, Shenzhen 518060, China*

ARTICLE INFO

Article history:
Received 25 September 2014
Received in revised form
12 December 2014
Accepted 19 December 2014
Available online 22 January 2015

Keywords:
Penalty function
Likelihood ratio method
Measurement error
SCAD
Single-index model

ABSTRACT

Adaptive testing for the partially linear single-index model (PLSIM) with error-prone linear covariates is considered. This is a fundamentally important and interesting problem for the current model because existing literature often assumes that the model structure is known before making inferences. In practice, this may result in an incorrect inference on the PLSIM. In this study, we explore whether the link function satisfies some special shape constraints by using an efficient penalized estimating method. For this we propose a model structure selection method by constructing a new testing statistic in the current setting with measurement error, which may enhance the flexibility and predictive power of this model under the case that one can correctly choose an adaptive shape and model structure. The finite sample performance of the proposed methodology is investigated by using some simulation studies and a real example from the Framingham Heart Study.

* Corresponding author at: School of Science, Nanjing University of Science and Technology, Nanjing 210094, Jiangsu, PR China.
E-mail address: stahzs@126.com (Z. Huang).

http://dx.doi.org/10.1016/j.stamet.2014.12.002
1572-3127/

1. Introduction

A partially linear single-index model with error-prone linear covariates can be described as

$$Y = g(\alpha_0^T X) + \beta_0^T Z + \varepsilon, \tag{1}$$

where ε is independent of $(X^T, Z^T)^T$ and has $E(\varepsilon) = 0$ and $\mathrm{Var}(\varepsilon) = \sigma^2$, $Y \in R$ is a response variable, X is a p-dimensional covariate vector, Z is a q-dimensional vector of completely unobserved latent covariates that is measured in an error-prone way; We write $Z = (\xi^T, W^T)^T$, and the covariates ξ is a $q_1 \times 1$ vector that cannot be observed directly and be generated from other auxiliary vector. That is to say, we may estimate it by using some auxiliary information, i.e. we assume that ξ is related to the observed variables η and V through $\xi = E(\eta|V)$, W is a vector of the remaining observed components. $g(\cdot)$ is an unknown link function, β_0 is a q-dimensional vector of unknown parameters, α_0 is a p-dimensional vector of unknown parameters and for model identifiability purpose, generally one may assume that $\|\alpha_0\| = 1$ and the first nonzero element of α_0 is positive, More details on the model identifiability can be found in Yu and Ruppert [13].

In the error-free case of model (1), the model (1) reduces to the partially linear single-index model (PLSIM), and various statistical inference methods for PLSIM have been studied in the literature. For example, Carroll et al. [1] considered a generalized version of PLSIM. Yu and Ruppert [13] proposed the spline smoothing method to estimate PLSIM. To further improve the efficiency of the existing estimation method, Xia and Härdle [12] suggested to use the minimum average variance estimation method to define an efficient and constructible estimators of the unknowns in PLSIM. Zhu and Xue [14] extended the empirical likelihood method to the PLSIM and constructed the confidence regions of the parametric components. Wang et al. [11] proposed a two-stage estimation procedure and claimed that their estimator improved those in the existing literature. Liang et al. [9] considered a penalized least-squared estimation method for PLSIM.

Huang [7] firstly proposed model (1) and employed the local linear smoothing method to consider the estimation of the unknown components. However, few attempts have been made to consider the model structure inferences of the current model. Our motivation of the current work is from the analysis of the well-known Framingham Heart Study. In this data set, it is generally accepted that measurement error exists on the cholesterol level because we cannot have the true cholesterol level but just the observed cholesterol level. Our main purpose is to check the effects of the standardized serum cholesterol, age and the binary variable smoking status on the systolic blood pressure. Our concern is focused on analyzing the Framingham Heart Study and selecting the model structure of interest from model (1), i.e. one may be interested in exploring whether the link function satisfies some special shape constraints. More details are provided in Section 3.

In this paper, firstly we extend the idea of nonconcave penalized least-squared method to model (1) to estimate the parametric components, which also has the nice theoretical properties as illustrated by Fan and Li [4]. Then, based on these given estimates of the parametric components and the local smoothing method, we construct a new testing statistic to find an adaptive semiparametric model structure. A new Wilks type of theorem is derived. Furthermore, the implementation of the algorithm for this step is also presented.

The remainder of the paper is organized as follows. In Section 2, we introduce the penalized estimators of the model (1). Further, the selection method for an adaptive semiparametric model structure is described and its asymptotic null distribution is derived. Simulated and real examples are analyzed in Section 3. The proofs of the main results are relegated to the Appendix.

2. Adaptive testing methodology

In this section, we will develop a new method to consider whether model (1) may admit an adaptive semiparametric structure. For this, a newly proposed test statistic is constructed and the large-sample property is further established.

In model (1), the unknown link function $g(\alpha_0^T X)$ may satisfy a shape constraint of interest. That is to say, the shape of the true curve $g(\alpha_0^T X)$ may be $g_0(X; \alpha_0)$, where $g_0(\cdot; \cdot)$ is a pre-specified function and α_0 is a vector of unknown parameters. Obviously, one can see that $g_0(X; \alpha_0)$ can have a more