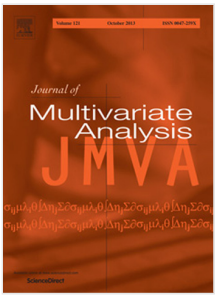

Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models

Yujie Li, Gaorong Li, Heng Lian, Tiejun Tong

PII: S0047-259X(16)30261-5
DOI: http://dx.doi.org/10.1016/j.jmva.2016.12.006
Reference: YJMVA 4200

To appear in: *Journal of Multivariate Analysis*

Received date: 23 March 2016

Please cite this article as: Y. Li, G. Li, H. Lian, T. Tong, Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models, *Journal of Multivariate Analysis* (2016), http://dx.doi.org/10.1016/j.jmva.2016.12.006

Profile forward regression screening for ultra-high dimensional semiparametric varying coefficient partially linear models

Yujie Li[a], Gaorong Li[*,b], Heng Lian[c], Tiejun Tong[d]

[a] *College of Applied Sciences, Beijing University of Technology, Beijing 100124, P. R. China*
[b] *Beijing Institute for Scientific and Engineering Computing, Beijing University of Technology, Beijing 100124, P. R. China*
[c] *Department of Mathematics, City University of Hong Kong, Hong Kong*
[d] *Department of Mathematics, Hong Kong Baptist University, Hong Kong*

Abstract

In this paper, we consider semiparametric varying coefficient partially linear models when the predictor variables of the linear part are ultra-high dimensional where the dimensionality grows exponentially with the sample size. We propose a profile forward regression (PFR) method to perform variable screening for ultra-high dimensional linear predictor variables. The proposed PFR algorithm can not only identify all relevant predictors consistently even for ultra-high semiparametric models including both nonparametric and parametric parts, but also possesses the screening consistency property. To determine whether or not to include the candidate predictor in the model of selected ones, we adopt an extended Bayesian information criterion (EBIC) to select the "best" candidate model. Simulation studies and a real data example are also carried out to assess the performance of the proposed method and to compare it with existing screening methods.

Key words: Varying coefficient partially linear model; profile forward regression; variable screening; screening consistency property; ultra-high dimension; EBIC

AMS 2010 subject classifications: primary 62G08; secondary 62J02

1. Introduction

In recent years, high-dimensional data analysis has become increasingly frequent and important in a large variety of areas such as health sciences, economics, finance, and epidemiology. The analysis of high-dimensional data poses many challenges for statisticians and thus calls for new statistical methodologies as well as theories; see Fan and Li [11].

To address these challenges, variable screening is an effective method of using a ranking criterion to select significant variables, particularly for statistical models with nonpolynomial dimensionality or "large p, small n" paradigms when p can be as large as an exponential of the sample size n; see Li et al. [24]. The main idea is to first apply a fast, reliable and efficient method to reduce the ultra-high dimensionality p from a large or huge

*Corresponding author. *E-mail address:* ligaorong@bjut.edu.cn (G. R. Li).
E-mail addresses: liyujie0207@emails.bjut.edu.cn (Yujie Li), henglian@cityu.edu.hk (Heng Lian), tongt@hkbu.edu.hk (Tiejun Tong)