Journal of Multivariate Analysis 126 (2014) 100–113

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

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

Efficient estimation for partially linear varying coefficient models when coefficient functions have different smoothing variables[☆]

Seong J. Yang, Byeong U. Park*

Université Catholique de Louvain, Belgium
Seoul National University, Republic of Korea

ARTICLE INFO

Article history:
Received 6 May 2013
Available online 25 January 2014

AMS 2000 subject classifications:
62G08
62G20

Keywords:
Partially linear varying coefficient models
Smooth backfitting
Semiparametric information bound
Profile likelihood

ABSTRACT

In this paper we consider partially linear varying coefficient models. We provide semiparametric efficient estimators of the parametric part as well as rate-optimal estimators of the nonparametric part. In our model, different nonparametric coefficients have different smoothing variables. This requires employing a projection technique to get proper estimators of the nonparametric coefficients, and thus conventional kernel smoothing cannot give semiparametric efficient estimators of the parametric components. We take the smooth backfitting approach in conjunction with the profiling technique to get semiparametric efficient estimators of the parametric part. We also show that our estimators of the nonparametric part achieve the univariate rate of convergence, regardless of the covariate's dimension. We report the finite sample properties of the semiparametric efficient estimators and compare them with those of other estimators.

1. Introduction

We consider the partially linear varying coefficient regression model (PLVCM) which takes the form

$$Y = \mathbf{X}^\top \boldsymbol{\beta} + \sum_{j=1}^{d} Z_j \alpha_j(U_j) + \varepsilon. \tag{1.1}$$

Here, Y is a response variable, $\mathbf{X} = (X_1, \ldots, X_p)^\top$, $\mathbf{Z} = (Z_1, \ldots, Z_d)^\top$ and $\mathbf{U} = (U_1, \ldots, U_d)^\top$ are covariates, ε is an error variable and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ is an unknown parameter vector. In this model the varying coefficients $\alpha_1, \ldots, \alpha_d$ are also unknown and have different 'smoothing variables' U_j. We write $\boldsymbol{\alpha}(\mathbf{U}) = (\alpha_1(U_1), \ldots, \alpha_d(U_d))^\top$. We let both the joint density of the covariate vector $(\mathbf{X}, \mathbf{Z}, \mathbf{U})$ and the density of the error term ε be unknown. Our goal is to present a semiparametric efficient estimator for $\boldsymbol{\beta}$. We also discuss the estimation of the nonparametric coefficients α_j.

The traditional linear regression model is very attractive due to its simplicity in estimation and interpretation. However, it may not be adequate in many complex situations of modern statistics. By adding a nonparametric part into the regression function, one can make the model more flexible while maintaining the simplicity of the linear model. To this end, many authors suggested various semiparametric models. Speckman [13] and Yu et al. [17] considered partially linear additive models. For partially linear additive models, the nonparametric part is added as a sum of univariate functions $\sum_{j=1}^{d} \alpha_j(U_j)$.

☆ This research was supported by the NRF Grant funded by the Korea government (MEST) No. 2010-0017437.
* Corresponding author at: Seoul National University, Republic of Korea.
E-mail addresses: bupark@stats.snu.ac.kr, bupark2000@gmail.com (B.U. Park).

0047-259X/$ – see front matter
http://dx.doi.org/10.1016/j.jmva.2014.01.004

Comparing this and the PLVCM at (1.1), the latter takes into account nonlinear interaction effects among covariates while the former does not.

Zhang et al. [18], Fan and Huang [6], and Ahmad et al. [1] studied a semiparametric model that is similar to (1.1). In their model, all coefficient functions α_j have a common smoothing variable, say $\mathbf{U}$ (or U in case it is univariate). We call this model PLVCM-1. It is well known that PLVCM-1 suffers from the curse of dimensionality when the common covariate $\mathbf{U}$ is of high-dimension. For this reason most studies have been focused on the case where the common smoothing variable is univariate, see [18,6,15,11,20], among others. Assuming a common univariate smoothing variable across all coefficient functions, one can only consider interaction effects among covariates in a very limited way. By allowing different coefficient functions to have different smoothing variables as in PLVCM at (1.1), one can accommodate more interaction effects. Here, we emphasize that fitting the nonparametric part of PLVCM is completely different from fitting that of PLVCM-1. While the conventional kernel smoothing works for PLVCM-1, it does not give proper estimators for PLVCM since it produces multivariate functions as estimators of the true α_j, see the discussion in Section 1 of [9]. For PLVCM, it requires a relevant 'projection' technique to get proper estimators of α_j, which makes the theory of semiparametric efficient estimation of the true $\boldsymbol{\beta}$ quite different from, and much more challenging than the one for PLVCM-1.

Varying coefficient models without the linear effects $\mathbf{X}^\top\boldsymbol{\beta}$ have been studied extensively since [8]. Most of them are for the case where the coefficient functions α_j have a common smoothing variable $\mathbf{U}$ or U. Examples include [5,7,3,4]. For this type of models, the conventional kernel smoothing works, and there is no need of a projection technique. It suffers from the curse of dimensionality, however, when $\mathbf{U}$ is of high-dimension. Recently, Yang et al. [16] studied the structured varying coefficient model where different coefficient functions α_j have different smoothing variables U_j as in the model (1.1). They applied the marginal integration approach as a projection technique. The method does not overcome the curse of dimensionality for a dimension $d \geq 5$, even with the structure in the model, which is typical with the marginal integration method. Later, Lee et al. [9] proposed a smooth backfitting technique that turns out to avoid the curse of dimensionality completely for all dimension, and [10] further extended the study to a fairly general varying coefficient model.

In the next section, we derive the semiparametric information bound for the estimation of the true $\boldsymbol{\beta}$ in the model (1.1). Roughly speaking, it refers to the minimal Fisher information among those for all 'regular' parametric submodels of (1.1). We show that the semiparametric information bound for PLVCM is greater than the one for PLVCM-1, meaning that one can estimate the true $\boldsymbol{\beta}$ more accurately in PLVCM than in PLVCM-1. In Section 3, we discuss semiparametric efficient estimation of the true $\boldsymbol{\beta}$ that achieves the minimal asymptotic variance among all 'regular' estimators of the true $\boldsymbol{\beta}$. The efficient estimation requires an initial $\sqrt{n}$-consistent estimator of the true $\boldsymbol{\beta}$, for a construction of which we take the profiling technique of Severini and Wong [12] and the smooth backfitting approach of Lee et al. [9]. We also show that our estimators of the true α_j enjoy the univariate optimal rate of convergence regardless of the dimension d. Some numerical evidences that support our theoretical findings are given in Section 4. All technical details are provided in the Appendix.

2. Semiparametric Fisher information

We present the semiparametric information bound for the parameter $\boldsymbol{\beta}$ in the PLVCM at (1.1). Let $\boldsymbol{\beta}_0$ denote the true coefficient vector of the regressor $\mathbf{X}$, $\boldsymbol{\alpha}_0$ the true coefficient function vector of the regressor $\mathbf{Z}$, and g_0 the true density function of the error term ε, in the model (1.1). We assume that ε is independent of $(\mathbf{X}, \mathbf{Z}, \mathbf{U})$. The true density belongs to the class of all symmetric densities g that are absolutely continuous with respect to the Lebesgue measure, have a derivative g' and finite Fisher information $\int (g')^2/g < \infty$.

To give an idea of the semiparametric information bound at the true parameter value $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, g_0)$, let $\boldsymbol{\beta} \mapsto (\boldsymbol{\alpha}_{\boldsymbol{\beta}}, g_{\boldsymbol{\beta}})$ be an arbitrary smooth mapping that passes through $(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, g_0)$. For each smooth mapping, there corresponds a parametric submodel $\{(\boldsymbol{\beta}, \boldsymbol{\alpha}_{\boldsymbol{\beta}}, g_{\boldsymbol{\beta}}) : \boldsymbol{\beta} \in \mathbb{R}^p\}$. One can obtain the Fisher information for each parametric submodel. Then, the semiparametric information bound equals the minimum of the parametric Fisher information bounds over all parametric submodels, see [2] for a general theory of semiparametric efficiency.

To implement the above idea, define for $1 \leq j \leq p$

$$\boldsymbol{\delta}_j = \frac{\partial}{\partial \beta_j}\boldsymbol{\alpha}_{\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}, \qquad \gamma_j = \frac{\partial}{\partial \beta_j}\log g_{\boldsymbol{\beta}}\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0}.$$

Let $\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}, g; y, \mathbf{x}, \mathbf{z}, \mathbf{u}) = \log g(y - \mathbf{x}^\top\boldsymbol{\beta} - \mathbf{z}^\top\boldsymbol{\alpha}(\mathbf{u}))$ denote the log-likelihood function at the observation $(y, \mathbf{x}, \mathbf{z}, \mathbf{u})$. Here, we omitted the logarithm of the density of $(\mathbf{X}, \mathbf{Z}, \mathbf{U})$ since it does not involve $(\boldsymbol{\beta}, \boldsymbol{\alpha})$. The score function for β_j under a submodel $\{(\boldsymbol{\beta}, \boldsymbol{\alpha}_{\boldsymbol{\beta}}, g_{\boldsymbol{\beta}}) : \boldsymbol{\beta} \in \mathbb{R}^p\}$ with a tangent vector $(\boldsymbol{\delta}_j, \gamma_j : 1 \leq j \leq p)$ is then given by

$$\begin{aligned}
\frac{\partial}{\partial \beta_j}\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}_{\boldsymbol{\beta}}, g_{\boldsymbol{\beta}})\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} &= \frac{\partial}{\partial \beta_j}\ell(\boldsymbol{\beta}, \boldsymbol{\alpha}_0, g_0)\bigg|_{\boldsymbol{\beta}=\boldsymbol{\beta}_0} + \frac{\partial}{\partial \boldsymbol{\alpha}} l(\boldsymbol{\beta}_0, \boldsymbol{\alpha}, g_0)\bigg|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}(\boldsymbol{\delta}_j) + \frac{\partial}{\partial \log g}\ell(\boldsymbol{\beta}_0, \boldsymbol{\alpha}_0, g)\bigg|_{g=g_0}(\gamma_j) \\
&= -x_j\frac{g_0'}{g_0}(\epsilon) - \mathbf{z}^\top\boldsymbol{\delta}_j(\mathbf{u})\frac{g_0'}{g_0}(\epsilon) + \gamma_j(\epsilon), \qquad (2.1)
\end{aligned}$$

where $\epsilon = y - \mathbf{x}^\top\boldsymbol{\beta}_0 - \mathbf{z}^\top\boldsymbol{\alpha}_0(\mathbf{u})$ and $\partial\ell/\partial\boldsymbol{\alpha}|_{\boldsymbol{\alpha}=\boldsymbol{\alpha}_0}(\boldsymbol{\delta}_j)$ denotes the Fréchet differential of ℓ at $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$ to the direction $\boldsymbol{\delta}_j$. Write $\Delta = (-\boldsymbol{\delta}_1, \ldots, -\boldsymbol{\delta}_p)$ and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)^\top$. Then, the Fisher information matrix $I(\Delta, \boldsymbol{\gamma})$ for this parametric submodel is