Published in final edited form as:
J Multivar Anal. 2008 September 1; 99(8): 1610–1634. doi:10.1016/j.jmva.2008.01.013.

A Two-Stage Approach for Semilinear In-Slide Models

Jinhong You [Postdoctoral Fellow] and **Haibo Zhou [Associate Professor]**
Department of Biostatistics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599-7400, USA
Jinhong You: jyou@bios.unc.edu; Haibo Zhou: zhou@bios.unc.edu

Abstract

The semilinear in-slide models (SLIMs) have been shown to be effective method for normalizing microarray data (Fan, *et al*. 2004). Using a backfitting method, Fan, Peng and Huang (2005) proposed a profile least squares (PLS) estimation for the parametric and nonparametric components. The general asymptotic properties for their estimator is not developed. In this paper, we consider a new approach, two-stage estimation, which enables us to establish the asymptotic normalities for both of the parametric and nonparametric component estimators. We further propose a plug-in bandwidth selector using the asymptotic normality of the nonparametric component estimator. The proposed method allow for the modeling of the aggregated SLIMs case where we can explicitly show that taking the aggregated information into account can improve both of the parametric and nonparametric component estimator by the proposed two-stage approach. Some simulation studies are conducted to illustrate the finite sample performance of the proposed procedures.

Key words and phrases

Semilinear regression; In-slide model; Two-stage estimation; Asymptotic normality; Aggregated information

1 Introduction

Microarray technology is an important tool for quantitatively monitoring gene expression patterns and has been widely used in functional genomics (see e.g. Schena *et al.*, 1995; Brown and Botstein 1999). Since great variations in experimental conditions exist in the microarray process it is essential to normalize the raw microarray data before any meaningful inference or analysis can be done. Useful normalization techniques developed include the global normalization method (e.g. Kroll and Wölfl 2002), the “lowess” method (e.g. Dudoit *et al.* 2002), the rank based procedure (e.g. Tseng *et al.* 2001). However, some restrictive biological assumptions are generally needed for normalization techniques. For example, the global normalization method needs an assumption that there is no print-tip block effect and no intensity effect. Without such an assumption, the global normalization method would be statistically biased. The “lowess” method requires an assumption that the average expression levels of up-and down-regulated genes at each intensity level are about the same in each print-

AMS 1980 classifications: primary 62H12; secondary 62A10

NIH-PA Author Manuscript
NIH-PA Author Manuscript

tip block. The rank based procedure assumes that there are not many genes that are up-regulated (or down-regulated).

New statistical approaches have been sought to relax those restrictive biological assumptions. For example, two-way semilinear models have been proposed to normalize the microarray data (Huang, *et al.* 2003, Huang and Zhang 2003, Huang, Wang and Zhang 2005). This method does not make the usual assumptions underlying the existing methods mentioned above. The two-way semilinear model approach can also incorporate uncertainty due to normalization into significant analysis of microarrays.

Fan, *et al.* (2004) proposed a method to estimate the intensity and print-tip effects by aggregating information from the replications in a microarray. Let G be the number of genes, I_g be the number of replications of the gth gene, R_{gi} and G_{gi} be the red (Cy5) and green (Cy3) intensities of the gth gene in the ith replication, respectively. Further, let Y_{gi} be the log-intensity ratio of red over green channels of the gth gene in the ith repetition, and let U_{gi} be the corresponding average of the log-intensities of the red and green channels. That is, $Y_{gi} = \log_2 R_{gi}/G_{gi}, U_{gi} = 1/2 \log_2(R_{gi}G_{gi})$. The following semilinear model was proposed by Fan, *et al.* (2004) to fit the intensity and print-tip block effects

$$Y_{gi}=\alpha_g+\beta_{r_{gi}}+\gamma_{c_{gi}}+m\left(U_{gi}\right)+\varepsilon_{gi}, \tag{1.1}$$

where α_g is the treatment effect associated with the gth gene, r_{gi} and c_{gi} are the row and column of print-tip block where the gth gene of the ith replication resides, β and γ are the row and column effects with constraints $\sum_{i=1}^{r}\beta_i=0$ and $\sum_{j=1}^{c}\gamma_j=0$, where r and c are the number of rows and columns of the print-tip blocks, $m(\cdot)$ is a smooth function of U representing the intensity effect, and ε_{gi}'s are random errors with mean zero and variance σ^2.

Using matrix notation, model (1.1) can be re-written as

$$\boldsymbol{Y}=\mathbf{B}\boldsymbol{\alpha}+\mathbf{X}\boldsymbol{\beta}+\mathbf{M}+\varepsilon, \quad n=\sum_{g=1}^{G}I_g \tag{1.2}$$

where $\mathbf{Y} = (Y_1,\ldots, Y_n)^T$ is the response, $\mathbf{B} = \text{blockdiag}(\mathbf{1}_{I_1},\ldots, \mathbf{1}_{I_G})$ with $\mathbf{1}_{I_g}$ being a vector of length I_g and all elements 1, $\mathbf{X} = (\mathbf{X}_1,\ldots,\mathbf{X}_n)^T$ is an $n \times p$ design matrix with p being the sum of the numbers of row and column, $\alpha = (\alpha_1,\ldots, \alpha_G)^T$ is the effect of gene, $\boldsymbol{\beta} = (\beta_1,\ldots, \beta_r, \gamma_1,\ldots, \gamma_c)^T$ is the print-tip block effect, $\mathbf{M} = (m(U_1),\ldots,m(U_n))^T$ is the intensity effect and $\varepsilon = (\varepsilon_1,\ldots, \varepsilon_n)^T$ is the random error.

Model (1.2) can be viewed as an extension of the usual fixed-effects parametric model to the semiparametric context. Such fixed-effects model is an appropriate specification if one is interested in a specific set of subjects and it has been widely applied in econometric analysis. (e.g. for example, Lichtenberg 1988,Honoré 1994,Baltagi 1995,Entorf 1997).

For the case where $I_g \equiv I$, Baltagi and Li (2002) proposed difference-based series (DBS) estimators for $\boldsymbol{\beta}$ and $m(\cdot)$. They established the asymptotic normality of the former and derived the convergence rate of the latter. Fan, Peng and Huang (2005) proposed profile least squares (PLS) estimators for $\boldsymbol{\beta}$ and $m(\cdot)$ by combining the local linear, least squares and backfitting procedures. They established the asymptotic normality of the former and derived the upper boundary of the mean squares error of the latter. You, Zhou, and Zhou (2005) proposed semiparametric least squares (SLE) estimators for $\boldsymbol{\beta}$ and $m(\cdot)$ by series approximating the nonparametric component. For DBS, PLS and SLE estimators, it is not easy to establish the asymptotic normality of the nonparametric component estimators. The reason is that the DBS and SLE involve the series approximation and the PLS uses a backfitting procedure. This hinders the application of these estimators in practice as it is difficult to select bandwidth and inference on the nonparametric component. In addition, Baltagi and Li (2002) and You, Zhou, and Zhou (2005) only consider the non-aggregated model.

Real microarray data often has different replication numbers reported, i.e. I_g may not always be the same across different g. This structure may arise from the fact that different studies have different replication number or that within a same study, uncontrollable experimental conditions such as image corruption, array fabrication error, etc, may lead to different I_g for different g (Golub *et al.* 1999, Alizadeh *et al.* 2000, Hendenfalk *et al.* 2001, Nguyen *et al.* 2004). Extension of model (1.2) under unequal I_g cases is undeveloped.

In this paper, we describe a two-stage estimation procedure. In the first stage, the series approximating estimation is used to obtain the series estimates of the parametric and nonparametric components. In the second stage, we input the first-stage estimates and eliminate the nuisance parameters α_g by difference. This transforms model (1.2) into an ordinary semilinear regression model. We then propose an ordinary profile least squares estimation for the parametric and nonparametric components, respectively. The asymptotic normalities of the proposed estimators are established. In particular, we show that the estimator of the parametric component achieves the semiparametric efficiency bound. We extend the two-stage estimate to the aggregated SLIMs case. Using the PLS estimation the aggregated information can only be used to improve the parametric components (Fan, Peng and Huang 2005). We explicitly demonstrate that under our two-stage estimation, the aggregated information can be used to improve both of the parametric and nonparametric component estimates.

The layout of the remainder of this paper is as follows. In Section 2 we describe the proposed two-stage estimation. In Section 3 we derive the asymptotic properties of the two-stage estimators. Extending the two-stage estimation to the aggregated SLIMs case is considered in Section 4. Section 5 presents results from numerical studies. Section 6 concludes. All proofs of main results are relegated to the Appendix.

2 A Two-Stage Procedure

Throught out this paper we assume that $G \to \infty$ and $2 \le I_g \le c$ for some fixed constant c. The two-stage estimation is as follows. In the first stage, the series approximating technique is used to obtain the series estimates of the parametric and nonparametric components, respectively. In the second stage, the first-stage estimates are input to the second stage and by differencing, we eliminate the nuisance parameters α_g and transform model (1.2) into an ordinary semilinear regression model. The ordinary profile least squares and local polynomial estimates are then obtained for the parametric and nonparametric components, respectively.

Since $m(u)$ is a smooth function, it can be approximated by $\boldsymbol{\zeta}^T(u)\boldsymbol{\vartheta}$ where $\boldsymbol{\zeta}(u) = (\zeta_{k_n 1}(u), \ldots, \zeta_{k_n k_n}(u))^T$ is a vector of approximating functions, such as power series or B-splines, $\boldsymbol{\vartheta}$ is an unknown k_n-variate constant vector and k_n is a positive integer which is dependent on n. Thus, model (1.2) can be written as

$$\boldsymbol{Y} = \mathbf{B}\boldsymbol{\alpha} + \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Xi}\boldsymbol{\vartheta} + \varepsilon^*, \tag{2.1}$$

where $\boldsymbol{\Xi}$ is an $n \times k_n$ matrix with i-th row being $\zeta(U_i) = (\zeta_{k_n 1}(U_i), \ldots, \zeta_{k_n k_n}(U_i))^T$, $\varepsilon^* = \varepsilon + \mathbf{M} - \boldsymbol{\Xi}\boldsymbol{\vartheta}$ and $\mathbf{M} = (m(U_1), \ldots, m(U_n))^T$. Define $\mathbf{M_B} = \mathbf{I}_n - \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$. Then pre-multiplying (2.1) by $\mathbf{M_B}$ leads to

$$\mathbf{M}_{\mathbf{B}}\boldsymbol{Y}=\mathbf{M}_{\mathbf{B}}\mathbf{X}\boldsymbol{\beta}+\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}\boldsymbol{\vartheta}+\mathbf{M}_{\mathbf{B}}\varepsilon^{*}. \quad (2.2)$$

If we take $\mathbf{M}_{\mathbf{B}}\varepsilon^*$ as the residuals, model (2.2) is a version of the usual linear regression. By the usual "profile" or "partialing out" formula, the estimator of $\boldsymbol{\beta}$ can be written as

$$\tilde{\boldsymbol{\beta}}_n=(\mathbf{X}^T\mathbf{M}_{\mathbf{B}}\mathbf{M}_{\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}}\mathbf{M}_{\mathbf{B}}\mathbf{X})^{-1}\mathbf{X}^T\mathbf{M}_{\mathbf{B}}\mathbf{M}_{\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}}\mathbf{M}_{\mathbf{B}}\boldsymbol{Y}, \quad (2.3)$$

where $\mathbf{M}_{\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}} = \mathbf{I}_n - \mathbf{P}_{\mathbf{M}_{\mathbf{B}}}\boldsymbol{\Xi} = \mathbf{I}_n - \mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}\,(\Xi^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi})^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}$ and $\mathbf{A}^{-}$ denotes any generalized inverse of matrix $\mathbf{A}$. An estimator of $\boldsymbol{\vartheta}$ is

$$\tilde{\boldsymbol{\vartheta}}_n=(\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi})^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\left(\boldsymbol{Y}-\mathbf{X}\tilde{\boldsymbol{\beta}}_n\right).$$

Then an obvious estimator of $m(u)$ is $\tilde{m}_n(u) = \boldsymbol{\zeta}^T(u)\tilde{\boldsymbol{\vartheta}}_n$, which is a nonparametric projecting estimator. Same as You, Zhou and Zhou (2005) we can establish the asymptotic normality of $\tilde{\boldsymbol{\beta}}_n$. However, it is a great challenge to establish the asymptotic normality of $\tilde{m}_n(u)$. The lack of asymptotic normality of the nonparametric component estimator poses difficulties for bandwidth selections and hinders statistical inference. In the following we will propose two-stage estiamtors for both of the parametric and nonparametric components and establish the asymptotic normality for both of them.

For convenience, let $\iota(g,i)=\sum_{g1=1}^{g-1}I_{g_1}+i$ and $Q(g,i)=\left(I_g-1\right)^{-1}\sum_{i_1=1,i_1\neq i}^{I_g}(Y_{\iota(g,i_1)}-\mathbf{X}_{\iota(g,i_1)}^T\tilde{\boldsymbol{\beta}}_n-\tilde{m}_n(U_{\iota(g,i_1)}))$ with $g = 1,\ldots,G$ and $i = 1,\ldots, I_g$. If subtracting $Q(g, i)$ from two sides of model (1.2) we have

$$Y_{\iota(g,i)}-Q(g,i)=\mathbf{X}_{\iota(g,i_1)}^T\boldsymbol{\beta}+m(U_{\iota(g,i)})+\varepsilon_{\iota(g,i)}+\alpha_g-Q(g,i).$$

According to Lemma 1 and Lemma 2 in the appendix, we have

$$\alpha_g-Q(g,i)=\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}+O_P(k_n/\sqrt{n}+k_n^{-3/2}).$$

Therefore, if we denote $Y_{\iota(g,i)}^*=Y_{\iota(g,i)}-Q(g,i)$ we have

$$Y_{\iota(g,i)}^*=\mathbf{X}_{\iota(g,i)}^T\boldsymbol{\beta}+m(U_{\iota(g,i)})+\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_i)}+O_p(k_n/\sqrt{n}+k_n^{-3/2}). \quad (2.4)$$

It is easy to see that (2.4) is an ordinary semilinear regression model. The ordinary profile least squares and local polynomial estimations can be used to estimate $\boldsymbol{\beta}$ and $m(\cdot)$. The detail is as follows. For any given $\boldsymbol{\beta}$, (2.4) can be written as

$$Y_{\iota(g,i)}^*-\mathbf{X}_{\iota(g,i)}^T\boldsymbol{\beta}=m(U_{\iota(g,i)})+\varepsilon_{\iota(g,i)}^{**},\quad g=1,\ldots,G,\ i=1,\ldots,I_g \quad (2.5)$$

where $\varepsilon^{**}_{\iota(g,i)}=\varepsilon_{\iota(g,i)}+\alpha_g-Q(g,i)$. This transforms the semilinear regression model into the usual nonparametric model. Now, apply a local linear regression technique in a small neighborhood of u_0, one can approximate $m(u)$ locally by a linear function

$$m(u)\approx m(u_0)+m'(u_0)(u-u_0)\equiv a+b(u-u_0)$$

with $m'(u)=\partial m/\partial u$. This leads to the following weighted local least squares problem: find a, b to minimize

$$\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{Y^{*}_{\iota(g,i)}-\mathbf{X}^T_{\iota(g,i)}\boldsymbol{\beta}-a-b(U_{\iota(g,i)}-u_0)\right\}^2K_h(U_{\iota(g,i)}-u_0), \tag{2.6}$$

where $K(\cdot)$ is a kernel function, h is a bandwidth and $K_h(\cdot)=K(\cdot/h)/h$. The solution to minimizing the sum in (2.6) is given by

$$(\widehat{a}(u),h\widehat{b}(u))^T=(\mathbf{D}^T_u\mathbf{W}_u\mathbf{D}_u)^{-1}\mathbf{D}^T_u\mathbf{W}_u(\boldsymbol{Y}^*-\mathbf{X}\boldsymbol{\beta}), \tag{2.7}$$

where

$$\mathbf{X}=\begin{pmatrix}\mathbf{X}^T_1\\ \vdots\\ \mathbf{X}^T_n\end{pmatrix}=\begin{pmatrix}X_{11}&\dots&X_{1p}\\ \vdots&\ddots&\vdots\\ X_{n1}&\dots&X_{np}\end{pmatrix},\quad \mathbf{D}_u=\begin{pmatrix}1&(U_1-u)/h\\ \vdots&\vdots\\ 1&(U_n-u)/h\end{pmatrix},$$

and

$$\mathbf{W}_u=\mathrm{diag}(K_h(U_1-u),\dots,K_h(U_n-u)).$$

Replacing $m(\cdot)$ by $\hat{a}(\cdot)$ in (2.5) results the following model

$$\widehat{Y}^{*}_{\iota(g,i)}=\widehat{\mathbf{X}}^T_{\iota(g,i)}\boldsymbol{\beta}+\varepsilon^{***}_{\iota(g,i)},\quad g=1,\dots,G \text{ and } i=1,\dots,I_g, \tag{2.8}$$

where

$$\begin{aligned}\widehat{Y}^{*}_{\iota(g,i)}&=Y^{*}_{\iota(g,i)}-(1,0)(\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\mathbf{D}_{U_{\iota(g,i)}})^{-1}\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\boldsymbol{Y}^*,\\ \widehat{\mathbf{X}}_{\iota(g,i)}&=\mathbf{X}_{\iota(g,i)}-(1,0)(\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\mathbf{D}_{U_{\iota(g,i)}})^{-1}\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\mathbf{X}\end{aligned}$$

and $\varepsilon^{***}_{\iota(g,i)}=\varepsilon^{**}_{\iota(g,i)}+\overline{m}(U_{\iota(g,i)})-\overline{\varepsilon}^{**}_{\iota(g,i)}$ with

$$\begin{aligned}\overline{m}(U_{\iota(g,i)})&=m(U_{\iota(g,i)})-(1,0)(\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\mathbf{D}_{U_{\iota(g,i)}})^{-1}\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\mathbf{M},\\ \overline{\varepsilon}^{**}_{\iota(g,i)}&=(1,0)(\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\mathbf{D}_{U_{\iota(g,i)}})^{-1}\mathbf{D}^T_{U_{\iota(g,i)}}\mathbf{W}_{U_{\iota(g,i)}}\varepsilon^{**},\\ \mathbf{M}&=(m(U_1),\dots,m(U_n))^T\quad\text{and } \varepsilon^{**}=(\varepsilon^{**}_1,\dots,\varepsilon^{**}_n)^T.\end{aligned}$$

NIH-PA Author Manuscript
NIH-PA Author Manuscript

Take $\varepsilon^{***}_{\iota(g,i_1)}$ as residuals and apply the least squares method to (2.8), we obtain a two-stage estimator of $\boldsymbol{\beta}$ as

$$\widehat{\boldsymbol{\beta}}_n=(\widehat{\mathbf{X}}^T\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^T\widehat{\boldsymbol{Y}}^*, \tag{2.9}$$

where $\mathbf{I}_n$ is an $n \times n$ identity matrix,

$$\mathbf{S}=\begin{pmatrix} (1,0)(\mathbf{D}^T_{U_1}\mathbf{W}_{U_1}\mathbf{D}_{U_1})^{-1}\mathbf{D}^T_{U_1}\mathbf{W}_{U_1} \\ \vdots \\ (1,0)(\mathbf{D}^T_{U_n}\mathbf{W}_{U_n}\mathbf{D}_{U_n})^{-1}\mathbf{D}^T_{U_n}\mathbf{W}_{U_n} \end{pmatrix}, \quad \boldsymbol{Y}^*=\begin{pmatrix} Y^*_1 \\ \vdots \\ Y^*_n \end{pmatrix}, \quad \begin{array}{l}\widehat{\mathbf{X}}=(\mathbf{I}_n-\mathbf{S})\mathbf{X}, \\ \widehat{\boldsymbol{Y}}^*=(\mathbf{I}_n-\mathbf{S})\boldsymbol{Y}^*.\end{array}$$

Correspondingly, a two-stage estimator of $m(\cdot)$ is

$$\widehat{m}_n(u)=(1,0)(\mathbf{D}^T_u\mathbf{W}_u\mathbf{D}_u)^{-1}\mathbf{D}^T_u\mathbf{W}_u(\boldsymbol{Y}^*-\mathbf{X}\widehat{\boldsymbol{\beta}}_n). \tag{2.10}$$

The error variance $\sigma^2=\mathrm{Var}(\varepsilon^2_1)$ is the quantity that describes the noise level. Apart from the intrinsic interest as parameters of the model, its estimation is essential in constructing confidence regions, model-based tests, model selection procedures, signal-to-noise ratio determination, and so on. Therefore, it is also essential to estimate it. We propose an estimate of σ^2 as follows

$$\widehat{\sigma}^2_n=\frac{1}{n+\sum_{g=1}^{G} I_g/(I_g-1)}(\boldsymbol{Y}^*-\mathbf{X}\widehat{\boldsymbol{\beta}}_n-\widehat{\mathbf{M}})^T(\boldsymbol{Y}^*-\mathbf{X}\widehat{\boldsymbol{\beta}}_n-\widehat{\mathbf{M}}).$$

In the next section, we will establish the asymptotic properties of $\hat{\boldsymbol{\beta}}_n, \hat{m}_n(\cdot)$ and $\widehat{\sigma}^2_n$.

3 Asymptotic Normality of the Two-Stage Estimators

To present the asymptotic properties of $\hat{\boldsymbol{\beta}}_n, \hat{m}_n(\cdot)$ and $\widehat{\sigma}^2_n$, we make the following assumptions

Assumption 1

$(\mathbf{X}_i, U_i, \varepsilon_i)$ *are independent and identically distributed as* $(\mathbf{X}_1, U_1, \varepsilon_1)$.

Assumption 2

(i) For very k_n *there is a nonsingular matrix* $\mathbf{M}$ *such that for* $\mathbf{M}\zeta(u)$, *the smallest eigenvalue of* $E[\mathbf{M}(\zeta(U_1)-E\zeta(U_1))]^{\otimes 2}$ *is bounded away from zero uniformly in* k_n.

(ii) There is a sequence of constants $\delta_0(k_n)$ *satisfying* $\sup_{u\in\mathcal{U}}\|\mathbf{M}\zeta(u)\| \le \delta_0(k_n)$ *and* k_n *satisfies that* $(\delta_0(k_n))^2k_n/n \to 0$ *as* $n \to \infty$, *where* $\mathcal{U}$ *is the support of* U_1, *and for a matrix* $\mathbf{A}$, $\|\mathbf{A}\| = tr(\mathbf{A}\mathbf{A}^T)$ *denotes the Euclidean norm of* $\mathbf{A}$.

Assumption 3

(i) $m(u)$ *and* $h_j(u) = E(X_{j1}|U_1 = u)$ *are twice continuously differentiable on* $\mathcal{U}$ *where* $j = 1,\ldots, p$.

(ii) For $m(u)$ *or* $h_j(u)$, $j = 1,\dots, p$, *there exist* $\boldsymbol{\vartheta} = (\vartheta_1,\dots,\vartheta_{k_n})^T$, *such that* $\sup_{u\in u} |g(u) - \boldsymbol{\vartheta}^T \boldsymbol{\zeta}(u)| O(k_n^{-2})$ *with* $g(u) = m(u)$ *or* $h_j(u)$.

(iii) $k_n = c_k n^{4/15+\nu}$ *for some constant* c_k *satisfying* $0 < c_k < \infty$ *and some* ν *satisfying* $0 \le \nu < 1/30$.

Assumption 4

The function $K(\cdot)$ *is a symmetric density function with compact support.*

Assumption 5

$h = c_h n^{-1/5}$ *for some constant* c_h *satisfying* $0 < c_h < \infty$.

Remark 1—*Assumption 2 is a standard assumption being used in series estimation methods. Assumption 3 says that the uniform approximation error to the function shrinks at the rate* k_n^{-2}. *Assumption 2 and Assumption 3 are not the easiest conditions but it is known that many series functions satisfy these conditions, e.g. power series and spline. Assumption 4 and Assumption 5 are standard assumptions used in kernel or local polynomial estimations.*

Under the above assumptions, the following theorem provides the asymptotic properties of $\hat{\boldsymbol{\beta}}_n, \hat{m}_n(\cdot)$ and $\widehat{\sigma}_n^2$

Theorem 1

Suppose that Assumption 1 to Assumption 5 hold. Then it holds that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{D} N\left(0, \{\lim_{n\to\infty} \frac{1}{n} \sum_{g=1}^{G} I_g^2 (I_g - 1)^{-1}\} \sigma^2 \boldsymbol{\Sigma}^{-1}\right) \quad as\ n \to \infty$$

where $\boldsymbol{\Sigma} = E(\boldsymbol{\Pi}_1 \boldsymbol{\Pi}_1^T)$ *and* $\boldsymbol{\Pi}_1 = \mathbf{X}_1 - E(\mathbf{X}_1 | U_1)$.

Theorem 2

Suppose that Assumption 1 to Assumption 5 hold. Then it holds that

$$\sqrt{nh}\left[\widehat{m}_n(u) - m(u) - \frac{h^2}{2} \frac{\mu_2^2 - \mu_1 \mu_3}{\mu_2 - \mu_1^2} m''(u)\right] \xrightarrow{D} N(0, \zeta(u)) \quad as\ n \to \infty$$

provided that $p(u) \neq 0$, *where* $\mu_j = \int_{-\infty}^{\infty} u^j K(u) du$, $\nu_j = \int_{-\infty}^{\infty} u^j K^2(u) du$,

$$\zeta(u) = \frac{\left\{\lim_{n\to\infty} n^{-1} \sum_{g=1}^{G} I_g^2 / \ I_g - 1\right\} \ \sigma^2 (\alpha_0^2 \nu_0 + 2\alpha_0 \alpha_1 \nu_1 + \alpha_1^2 \nu_2)}{p(u)},$$

with $\alpha_0 = -\mu_2 / (\mu_2 - \mu_1^2)$ *and* $\alpha_1 = \mu_1 / (\mu_2 - \mu_1^2)$ *and* $p(\cdot)$ *is the density function of* U_1.

Remark 2—*According to Theorem 1, when* $I_g \equiv I$ *the asymptotic covariance matrix of* $\hat{\boldsymbol{\beta}}_n$ *reduces to* $I/(I-1)\sigma^2\Sigma^{-1}$, *i.e the semiparametric efficient boundary (Fan, Peng and Huang 2005).*

Theorem 3

Suppose that Assumption 1 to Assumption 5 hold. If $E\varepsilon_1^4 < \infty$ holds, then

$$\sqrt{n}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{D} N(0, \kappa) \quad as\ n \to \infty,$$

where $\kappa = \theta_1^0 E(\varepsilon_1^4)\theta_2^0\sigma^4$ *with* $\tau(n) = n / \left\{n + \sum_{g=1}^{G} I_g / \left(I_g - 1\right)^2\right\}$,

$$\theta_1^0 = \lim_{n\to\infty} \tau(n) \sum_{g=1}^{G} \left\{1 + \frac{1}{(I_g - 1)^2} + \frac{2}{(I_g - 1)}\right\} I_g$$

and

$$\theta_2^0 = \lim_{n\to\infty} \tau(n) \sum_{g=1}^{G} \frac{1}{(I_g - 1)^3}(-I_g^4 + 2I_g^3 + 6I_g^2 + I_g).$$

Further, we define

$$\widehat{\mathbf{\Sigma}}_n = \sum_{g=1}^{G} I_g^2 (I_g - 1)^{-1} \widehat{\sigma}_n^2 (\widehat{\mathbf{X}}^T \widehat{\mathbf{X}})^{-1},$$

$$\widehat{\psi}_{\iota(g,i)} = (Y_{\iota(g,i)} - \mathbf{X}_{\iota(g,i)}^T \widehat{\boldsymbol{\beta}}_n - \widehat{m}_n(U_{\iota(g,i)})) - (Y_{\iota(g,i-1)} - \mathbf{X}_{\iota(g,i-1)}^T \widehat{\boldsymbol{\beta}}_n - \widehat{m}_n(U_{\iota(g,i-1)}))$$

for $g = 1,\ldots,G$, $i = 2,\ldots, I_g$,

$$\theta_1 = \tau(n) \sum_{g=1}^{G} \left\{1 + \frac{1}{(I_g - 1)^2} + \frac{2}{(I_g - 1)}\right\} I_g,$$

$$\theta_2 = \tau(n) \sum_{g=1}^{G} \frac{1}{(I_g - 1)^3}(-I_g^4 + 2I_g^3 + 6I_g^2 + I_g), \quad \theta_3 = \sum_{g=1}^{G}(4I_g - 2),$$

$$\theta_4 = \sum_{g=1}^{G} \left\{(I_g - 1)(I_g + 2) + 4I_g\right\} \text{ and } \widehat{\kappa}_n = \theta_1/\theta_3 \sum_{g=1}^{G}\sum_{i=1}^{I_g} \widehat{\psi}_{\iota(g,i)}^4 + \{\theta_2 - (\theta_1\theta_4)/\theta_3\}\widehat{\sigma}_n^4$$

The next theorem shows that $\hat{\mathbf{\Sigma}}_n$ and $\hat{\kappa}_n$ are consistent estimators of $\lim_{n\to\infty} n^{-1} \sum_{g=1}^{G} I_g^2 \cdot \left(I_g - 1\right)^{-1} \sigma^2 \mathbf{\Sigma}$ and κ, respectively.

Theorem 4

Suppose that Assumption 1 to Assumption 5 hold. If $E\varepsilon_1^4 < \infty$ holds, then

$$\widehat{\mathbf{\Sigma}}_n \to_p \left\{\lim_{n\to\infty} n^{-1} \sum_{g=1}^{G} I_g^2 (I_g - 1)^{-1}\right\} \sigma^2 \mathbf{\Sigma} \text{ and } \widehat{\kappa}_n \to_p \kappa \quad as\ n \to \infty.$$

4 Two-stage Estimation for the Aggregated SLIM

In so far, the intensity effect and the gene effect were estimated by using the information within one slide. Therefore, the arrays are allowed to have different gene effect, namely, α_g can be slide-dependent. When samples were drawn from different subjects this is reasonable. However, in many practical situations, the sample may come from the same subject. In those cases, it is natural to assume that the gene effects are the same across arrays and the information from other arrays can be aggregated. This assumption is helpful for improving the precision and for assessing the quality of an array using the coefficient of variation (Tseng, *et al.* 2001). Therefore, Fan, Peng and Huang (2005) further proposed an aggregated SLIM. This kind of aggregation idea is also appeared in the work of Huang, Wang and Zhang (2003) for a very different semiparametric model. The aggregated SLIM is defined as

$$Y_{ij}=\mathbf{B}_{ij}^T\boldsymbol{\alpha}+\mathbf{X}_{ij}^T\boldsymbol{\beta}_j+m_j(U_{ij})+\varepsilon_{ij},\ i=1,\dots,n,\ j=1,\dots,J. \tag{4.1}$$

where $\mathbf{Y}_j=(Y_{1j},\dots,Y_{nj})^T$, $\mathbf{B}_j=(\mathbf{B}_{1j},\dots,\mathbf{B}_{nj})^T$, $\mathbf{X}_j=(\mathbf{X}_{1j},\dots,\mathbf{X}_{nj})^T$, $\mathbf{U}_j=(U_{1j},\dots,U_{nj})^T$, $\boldsymbol{\alpha}=(\alpha_1,\dots,\alpha_G)^T$, $\boldsymbol{\beta}_j=(\beta_{1j},\dots,\beta_{p_j j})^T$ and $\varepsilon_j=(\varepsilon_{1j},\dots,\varepsilon_{nj})^T$.

Fan, Peng and Huang (2005) proposed an aggregated profile least squares (APLS) estimator for $\boldsymbol{\beta}=(\boldsymbol{\beta}_1^T,\dots,\boldsymbol{\beta}_J^T)^T$ and describe an estimation for the nonparametric components. We here propose an aggregated two-stage procedure.

4.1 Estimating the parametric component

We will investigate two cases. One is that $\mathbf{X}_{ij1}$ and $\mathbf{X}_{ij2}$ are independent and the other is $\mathbf{X}_{ij1}$ and $\mathbf{X}_{ij2}$ are dependent, where $j_1 \neq j_2$.

Case 1—Suppose that $\tilde{\boldsymbol{\beta}}_{jn}$ and $\tilde{m}_{jn}(\cdot)$ are series estimators of $\boldsymbol{\beta}_j$ and $m_j(\cdot)$, respectively which are based on individual equation. Let

$$\nabla_{\iota(g,i),1}=Y_{\iota(g,i),j}-\mathbf{X}_{\iota(g,i),j}^T\tilde{\boldsymbol{\beta}}_{jn}-\tilde{m}_{jn}(U_{\iota(g,i),j}).$$

For fixed j, if subtracting

$$(I_gJ-1)^{-1}\left\{\sum_{i_1=1,i_1\neq i}^{I_g}\nabla_{\iota(g,i_1),j}+\sum_{j_1=1,j_1\neq j}^{J}\sum_{i_1=1}^{I_g}\nabla_{\iota(g,i_1),j_1}\right\}$$

from the two sides of model (4.1) we have

$$\begin{aligned}
&Y_{\iota(g,i),j}-\frac{1}{I_gJ-1}\left\{\sum_{i_1=1,i_1\neq i}^{I_g}\nabla_{\iota(g,i_1),j}+\sum_{j_1=1,j_1\neq j}^{J}\sum_{i_1=1}^{I_g}\nabla_{\iota(g,i_1),j_1}\right\}\\
=\ &\mathbf{X}_{\iota(g,i),j}^T\boldsymbol{\beta}_j+m(U_{\iota(g,i),j})+\varepsilon_{\iota(g,i),j}+\alpha_g-\frac{1}{I_gJ-1}\left\{\sum_{i_1=1,i_1\neq i}^{I_g}\nabla_{\iota(g,i_1),j}+\sum_{j_1=1,j_1\neq j}^{J}\sum_{i_1=1}^{I_g}\nabla_{\iota(g,i_1),j_1}\right\}\\
=\ &\mathbf{X}_{\iota(g,i),j}^T\boldsymbol{\beta}_j+m(U_{\iota(g,i),j})+\varepsilon_{\iota(g,i),j}-\frac{1}{I_gJ-1}\left\{\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1),j}+\sum_{j_1=1,j_1\neq j}^{J}\sum_{i_1=1}^{I_g}\varepsilon_{\iota(g,i_1),j}\right\}\\
&+O_p(\max_{1\le j\le J}k_{jn}/\sqrt{n}+\max_{1\le j\le J}k_{jn}^{-3/2}).
\end{aligned}$$

Therefore, applying the usual profile least squares estimation we can obtain an aggregated two-stage estimator of $\boldsymbol{\beta}_j$ as

$$\widehat{\boldsymbol{\beta}}_{jn}^{(1)A} = (\widehat{\mathbf{X}}_j^T \widehat{\mathbf{X}}_j)^{-1} \widehat{\mathbf{X}}_j^T \widehat{\boldsymbol{Y}}_j^{(1)*},$$

where $\mathbf{S}_j$, $\hat{\mathbf{X}}_j$ have the same definitions as $\mathbf{S}$ and $\hat{\mathbf{X}}$, the $\iota(g, i)$th element of $\boldsymbol{Y}_j^{(1)*}$ is $Y_{\iota(g,i),j} - \left(I_g J - 1\right)^{-1} \left\{ \sum_{i_1=1, i_1 \neq i}^{I_g} \nabla_{\iota(g,i_1),j} + \sum_{j_1=1, j_1 \neq j}^{J} \sum_{i_1=1}^{I_g} \nabla_{\iota(g,i_1),j_1} \right\}$ and $\widehat{\boldsymbol{Y}}_j^{(1)*} = \left(\mathbf{I}_n - \mathbf{S}_j\right) \boldsymbol{Y}_j^{(1)*}$.

Case 2—For fixed j, if subtracting $\left\{I_g (J-1)\right\}^{-1} \sum_{j_1=1}^{J} \sum_{i_1=1, i_1 \neq i}^{I_g} \nabla_{\iota(g,i_1),j_1}$ from the two sides of model (4.1) we have

$$Y_{\iota(g,i),j} - \frac{1}{I_g(J-1)} \sum_{j_1=1}^{J} \sum_{i_1=1, i_1 \neq i}^{I_g} \nabla_{\iota(g,i_1),j_1} = \mathbf{X}_{\iota(g,i),j}^T \boldsymbol{\beta}_j + m(U_{\iota(g,i),j})$$
$$+ \varepsilon_{\iota(g,i),j} - \frac{1}{I_g(J-1)} \sum_{j_1=1}^{J} \sum_{i_1=1, i_1 \neq i}^{I_g} \varepsilon_{\iota(g,i_1),j_1} + O_p(\max_{1 \le j \le J} k_{jn} / \sqrt{n} + \max_{1 \le j \le J} k_{jn}^{-3/2}).$$

Therefore, applying the usual profile least squares estimation we can obtain an aggregated two-stage estimator of $\boldsymbol{\beta}_j$ as

$$\widehat{\boldsymbol{\beta}}_{jn}^{(2)A} = (\widehat{\mathbf{X}}_j^T \widehat{\mathbf{X}}_j)^{-1} (\widehat{\mathbf{X}}_j^T \widehat{\boldsymbol{Y}}_j^{(2)*},$$

where the $\iota(g, i)$th element of $\boldsymbol{Y}_j^{(2)*}$ is $Y_{\iota(g,i),j} - \left\{I_g (J-1)\right\}^{-1} \sum_{j_1=1}^{J} \sum_{i_1=1, i_1 \neq i}^{I_g} \nabla_{\iota(g,i_1),j_1}$.

For $\widehat{\boldsymbol{\beta}}_{jn}^{(1)A}$ and $\widehat{\boldsymbol{\beta}}_{jn}^{(2)A}$ we have the following asymptotic properties.

Theorem 5—*Under some regularity conditions (same as Assumption 1 to Assumption 5) it holds that*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{jn}^{(1)A} - \boldsymbol{\beta}_j) \underset{\rightarrow}{D} N\left(0, \lim_{n \to \infty} n^{-1} \sum_{g=1}^{G} \frac{J I_g^2}{J I_g - 1} \sigma^2 \sum\nolimits_j^{-1}\right) \text{ as } n \to \infty$$

where $\sum_j = E(\boldsymbol{\Pi}_{1j} \boldsymbol{\Pi}_{1j}^T)$ *and* $\boldsymbol{\Pi}_{1j} = \mathbf{X}_{1j} - E\left(\mathbf{X}_{1j} \middle| U_{1j}\right)$.

Theorem 6—*Under some regularity conditions (same as Assumption 1 to Assumption 5) it holds that*

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{jn}^{(2)A} - \boldsymbol{\beta}_j) \overset{D}{\rightarrow} N\left(0, \lim_{n \to \infty} n^{-1} \sum_{g=1}^{G} \frac{\left\{J\left(I_g - 1\right) + 1\right\}}{J\left(I_g - 1\right)} \sigma^2 \sum\nolimits_j^{-1}\right) \quad \text{as } n \to \infty$$

where $\boldsymbol{\Sigma}_j$ *is defined in Theorem 5.*

Remark 3: *From Theorem 5 and 6 we can see the aggregated information can be used to improve the two-stage estimators for the parametric components and the degree of improvement depend on* $\mathbf{X}_{ij1}$ *and* $\mathbf{X}_{ij2}$ *being independent or dependent. Moreover, when* $I_g \equiv I,\ \lim_{n\to\infty} n^{-1}\sum_{g=1}^{G}\frac{JI_g^2}{JI_g-1}$ *reduces to JI/(JI − 1). Thus, according to Fan, Peng and Huang (2005), our aggregated two-stage estimator has the same asymptotic covariance as that of the aggregated PLS estimator*.

4.2 Estimating the nonparametric components

We propose an aggregated local linear estimator of $m_j(\cdot)$ for Case 1 and 2. In Case 1, it has the form

$$\widehat{m}_{jn}^{(1)A}(u) = (1,0)\,(\mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{D}_{ju})^{-1} \mathbf{D}_{ju}^T \mathbf{W}_{ju} (\mathbf{Y}_j^{(1)*} - \mathbf{X}_j \widehat{\boldsymbol{\beta}}_{jn}^{(1)A}).$$

In Case 2, it has the form

$$\widehat{m}_{jn}^{(2)A}(u) = (1,0)\,(\mathbf{D}_{ju}^T \mathbf{W}_{ju} \mathbf{D}_{ju})^{-1} \mathbf{D}_{ju}^T \mathbf{W}_{ju} (\mathbf{Y}_j^{(1)*} - \mathbf{X}_j \widehat{\boldsymbol{\beta}}_{jn}^{(2)A}).$$

For $\widehat{m}_{jn}^{(1)A}(u)$ and $\widehat{m}_{jn}^{(2)A}(u)$, we have the following asymptotic properties.

Theorem 7— *Under some regularity conditions (same as Assumption 1 to Assumption 5) it holds that*

$$\widehat{m}_{jn}^{(1)A}(u) - \widehat{m}_{jn}^{(2)A}(u) = o_p\left\{h^2 + \frac{1}{\sqrt{nh}}\right\}.$$

Further,

$$\sqrt{nh}\quad\left[\widehat{m}_{jn}^{(1)A}(u) - m_j(u) - \frac{h^2}{2}\frac{\mu_2^2 - \mu_1\mu_3}{\mu_2 - \mu_1^2} m_j''(u)\right] \xrightarrow{D} N(0, \zeta_j^A(u)) \quad as \quad n \to \infty$$

provided that $p_j(u) \neq 0$, *where*

$$\zeta_j^A(u) = \frac{\{\lim_{n\to\infty} n^{-1}\sum_{g=1}^{G}\frac{JI_g^2}{JI_g-1}\}\sigma^2(\alpha_0^2\nu_0 + 2\alpha_0\alpha_1\nu_1 + \alpha_1^2\nu_2)}{p_j(u)},$$

and $p_j(u)$ *is the density function of* U_{1j}.

Remark 4— *From Theorem 7, we can see that taking the aggregated information into account can improve the estimate of the nonparametric component as well*.

5 Simulation Studies

In this section, we conduct some simulations to show the finite sample performance of the estimators in last sections. In order to compare our estimators with those in Fan, Peng and Huang (2005) we take Example 1 of Fan, Peng and Huang (2005).

Example 1

We select G = 100, 200, 400, 800 and I = 2, 3, 4. For each pair of (G, I), we simulate 200 datasets from model (1.2). The details of simulation scheme for this example are as follows:

α_g: The expression levels of the genes are generated from the standard double-exponential distribution.

β: For the row effects, first generate $\{\beta_i', i=1, \dots, 4\}$ from $N(0, 0.5)$, then set $\beta_i=\beta_i'-\overline{\beta'}$, which will guarantee that $\sum_{i=1}^{4}\beta_i=0$. The column effects are generated in the same way.

U: The intensity is generated from a mixture distribution. We generate u from probability $0.0004(u-6)^3 I(6 < u < 16)$ with probability 0.7 and from uniform distribution over [6, 16] with probability 0.3.

$m(\cdot)$: Set the function $m(u)=\sqrt{5}(\sin(u)-0.2854)$, where expectation is 0.

X: For each given gene, its associated block is assigned at random at one of 32 print-tip blocks.

ε: ε_{gi} is generated from the standard normal distribution.

For the proposed estimation, in first stage, we use a cubic B-spline basis function defined by

$$\zeta(u|u^0, \dots, u^4)=\frac{1}{3!}\sum_{j=0}^{4}(-1)^j\begin{pmatrix} 4 \\ j \end{pmatrix}[\max(0, u-u^j)]^3,$$

where $u^0,\dots, u^4$ are the evenly-spaced design knots. In the second stage, we take the Gaussian kernel, i.e.

$$K_h(u)=\frac{1}{h\sqrt{2\pi}}\exp(-u^2/2h^2).$$

and the bandwidth is selected by plug-in method. The performance of the estimators is assessed by the mean squared errors (MSEs). The results are summarized in Table 1 and Figure 1.

From Table 1 and Figure we can see that the two-stage estimators almost has the same finite sample performance as that of the profile least squares estimators. This phenomena is also observed for the case of aggregation across arrays. We here omit the detail.

6 Concluding Remarks

In this paper, we have proposed a two-stage estimation procedure for the semilinear in-slide models. The main advantage of our approach over the existing ones is that we can establish the asymptotic normalities for the corresponding parametric and nonparametric component estimators, respectively. We further extended the two-stage estimation to aggregated semilinear in-slide models. The advantage of the two-stage estimation over the existed estimations in this case is that we can explicitly show that taking the aggregated information can lead to improvement in both the the parametric and nonparametric component estimators. The significance of developing these asymptotic normalities lies in that we can do bandwidth selection and statistical inference for the interested parametric and nonparametric components.

This is still an fast evolving area of research and additional effort in this direction is warranted. For example, how to take the heteroscedastic into account to improve the two-stage estimation is still an open problem.

Acknowledgments

This research is supported by a grant from National Institute of Health (CA 79949).

Appendix. Proof of Main Results

Lemma 1

Let $(X_1, Y_1),\ldots, (X_n, Y_n)$ be i.i.d random vectors, where the Y_i's are scalar random variables. Further assume that $E|Y_i|^4 < \infty$ and $\sup_x \int |y|^4 f(x, y)dy < \infty$, where f denotes the joint density of (X, Y). Let K be a bounded positive function with a bounded support, and satisfies Lipschitz's condition. Then if $nh^8 \to 0$ and $nh^2/(\log n)^2 \to \infty$, it holds that

$$\sup_X \left|\frac{1}{n}\sum_{i=1}^{n} [K_h(X_i - X)Y_i - E\{K_h(X_i - X)Y_i\}]\right| = O_p\left(\left\{\frac{\log\left(1/h\right)}{nh}\right\}^{\frac{1}{2}}\right).$$

.

The proof of Lemma 1 follows immediately from the result of Mack and Silverman (1982).

Lemma 2

Suppose that Assumption 3 to Assumption 5 hold. Then it holds that $\lim_{n\to\infty}\frac{1}{n}\widehat{\mathbf{X}}^T\widehat{\mathbf{X}}=\mathbf{\Sigma}$ where $\hat{\mathbf{X}}$ is defined in Section 2 and $\mathbf{\Sigma}$ is defined in Theorem 1.

The proof of Lemma 2 is trivial. We here omit the detail.

Lemma 3

Suppose that Assumption 1 to Assumption 3 hold. Then we have $\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = O_p(n^{-1/2})$ Further,

$$\tilde{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \left\{\mathbf{\Pi}^T(\mathbf{I}_n - \mathbf{P_B})\mathbf{\Pi}\right\}^{-1}\mathbf{\Pi}^T(\mathbf{I}_n - \mathbf{P_B})\varepsilon + o_p(n^{-\frac{1}{2}})$$

where $\mathbf{\Pi} = (\mathbf{\Pi}_1,\ldots, \mathbf{\Pi}_n)^T$, $\mathbf{\Pi}_i = \mathbf{X}_i - E(\mathbf{X}_i|U_i)$ *and* $\mathbf{P_B} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T$.

Lemma 4

Suppose that Assumption 1 to Assumption 3 hold. Then we have

a. $\lim_{n\to\infty} \|\tilde{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}\| \to_p 0$;

b. $\tilde{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}=O_p(k_n^{1/2}/n^{1/2}+k_n^{-2})$;

c. $\sup_{u\in u}\tilde{m}_n(u)-m(u)=O_p(k_n/\sqrt{n}+k_n^{-3/2})$;

Further,

d. $\tilde{\boldsymbol{\vartheta}}_n-\boldsymbol{\vartheta}=\left\{\boldsymbol{\zeta}^T(\mathbf{I}_n-\mathbf{P_B})\boldsymbol{\zeta}\right\}^{-1}\boldsymbol{\zeta}^T(\mathbf{I}_n-\mathbf{P_B})\varepsilon+\left\{\boldsymbol{\zeta}^T(\mathbf{I}_n-\mathbf{P_B})\boldsymbol{\zeta}\right\}^{-1}\boldsymbol{\zeta}^T(\mathbf{I}_n-\mathbf{P_B})(m(U_1)-\boldsymbol{\zeta}^T(U_1)\boldsymbol{\vartheta},\ldots,m(U_n)-\boldsymbol{\zeta}^T(U_n)\boldsymbol{\vartheta})+O_p(k_n^{3/2}/n+n^{-1/2})$.

The proof of Lemma 3 is same as that of Theorem 1 in You, Zhou and Zhou (2005). Applying the root-n consistency of $\tilde{\boldsymbol{\beta}}_n$, combining the proof of Theorem 1 in Horowitz and Mammen (2004) we can show Lemma 4 holds. We here omit the detail.

Proof of Theorem 1

For convenience, let $\Delta_g(-i)=\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(\mathbf{X}_{\iota(g,i_1)}^T(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)+m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))$ for $g=1,\ldots,G$. Then, according to the definition of $\hat{\boldsymbol{\beta}}_n$ it can be verified that

$$
\begin{aligned}
&\widehat{\boldsymbol{\beta}}_n-\boldsymbol{\beta}\\
=\;&(\widehat{\mathbf{X}}^T\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^T(\mathbf{I}_n-\mathbf{S})\Bigg\{\mathbf{X}\boldsymbol{\beta}+\mathbf{M}+\varepsilon-\Bigg(\tfrac{1}{I_1-1}\sum_{i_1=2}^{I_1}\varepsilon_\iota(1,i_1),\cdots,\tfrac{1}{I_1-1}\sum_{i_1=1,i_1\neq I_1}^{I_1}\varepsilon_\iota(1,i_1),\\
&\cdots,\tfrac{1}{I_G-1}\sum_{i_1=1,i_1\neq I_G}^{I_G}\varepsilon_\iota(G,i_1)\Bigg)^T\Bigg\}\\
&-(\widehat{\mathbf{X}}^T\widehat{\mathbf{X}})^{-1}\widehat{\mathbf{X}}^T(\mathbf{I}_n-\mathbf{S})(\Delta_1(-1),\ldots,\Delta_1(-I_1),\ldots,\Delta_G(-I_G))^T\\
=\;&(\widehat{\mathbf{X}}^T\widehat{\mathbf{X}})^{-1}J_1+(\widehat{\mathbf{X}}^T\widehat{\mathbf{X}})^{-1}J_2,\text{ say.}
\end{aligned}
$$

Therefore, combining Lemma 2 in order to complete the proof we just need to show that

$$
\frac{1}{\sqrt{n}}J_1\to_D N(0,\lim_{n\to\infty}\frac{1}{n}\sum_{g=1}^{G}\frac{I_g^2}{I_g-1}\sigma^2\boldsymbol{\Sigma})\quad\text{as } n\to\infty \tag{A.1}
$$

and $J_2=o_p(n^{1/2})$. Following the same argument as the proof of Theorem 1 in Fan and Huang (2005) we have

$$
\frac{1}{\sqrt{n}}J_1=\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\boldsymbol{\Pi}_{\iota(g,i)}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)+o_p(1).
$$

Since $\sum_{i=1}^{I_g}\boldsymbol{\Pi}_{\iota(g,i)}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)$'s are independent random variables with mean zero and finite covariance matrix

$$
\begin{aligned}
\mathrm{Cov}\left\{\sum_{i=1}^{I_g}\boldsymbol{\Pi}_{\iota(g,i)}\left(\varepsilon_{\iota(g,i)}-\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\right\}&=I_g\boldsymbol{\Sigma}\mathrm{Cov}\left\{\left(\varepsilon_{\iota(g,1)}-\tfrac{1}{I_g-1}\sum_{i_1=2}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\right\}\\
&=I_g^2/\left(I_g-1\right)\sigma^2\boldsymbol{\Sigma},
\end{aligned}
$$

by central limit theorem and Slustky's theorem (A.1) holds. Moreover,

$$\begin{aligned}&\tfrac{1}{n}J_2=\tfrac{1}{n}\widehat{\mathbf{X}}^T(\mathbf{I}_n-\mathbf{S})(\Delta_1(-1),\ldots,\Delta_1(-I_1),\ldots,\Delta_G(-I_G))^T\\ =\ &\tfrac{1}{n}\widehat{\mathbf{X}}^T(\Delta_1(-1),\ldots,\Delta_1(-I_1),\ldots,\Delta_G(-I_G))^T-\tfrac{1}{n}\widehat{\mathbf{X}}^T\mathbf{S}(\Delta_1(-1),\ldots,\Delta_1(-I_1),\ldots,\Delta_G(-I_G))^T\\ =\ &J_{21}+J_{22},\ \text{say}.\end{aligned}$$

Let $\mathbf{O}(u)=(1,0)(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\mathbf{D}_u^T\mathbf{W}_u$. By definition of $\mathbf{X}$^ it holds that

$$\begin{aligned}J_{21}=&\tfrac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(\mathbf{X}_{\iota(g,i_1)}^T(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n)+m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))\\ &+\tfrac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{O}(U_{\iota(g,i)})\mathbf{\Pi}\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(\mathbf{X}_{\iota(g,i_1)}^T\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)+m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))\\ &+\tfrac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{\mathbf{h}(U_{\iota(g,i)})-\mathbf{O}(U_{\iota(g,i)})\mathbf{H}\right\}\cdot\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(\mathbf{X}_{\iota(g,i_1)}^T\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)+m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))\\ =&J_{211}+J_{212}+J_{213},\ \text{say}\end{aligned}$$

where $\mathbf{h}(u)=(E(X_{11}|U_1=u),\ldots,E(X_{p1}|U_1=u))^T$ and $\mathbf{H}=(\mathbf{h}(U_1),\ldots,\mathbf{h}(U_n))^T$. By Fan and Huang (2005) it holds that

$$\max_{1\le g\le G}\max_{1\le i\le I_g}\|\mathbf{O}(U_{\iota(g,i)})\mathbf{\Pi}\|=O_p\left(h^2+\frac{1}{\sqrt{nh}}\right)$$

and

$$\max_{1\le g\le G}\max_{1\le i\le I_g}\|\mathbf{h}(U_{\iota(g,i)})-\mathbf{O}(U_{\iota(g,i)})\mathbf{H}\|=O_p\left(h^2+\frac{1}{\sqrt{nh}}\right).$$

Therefore, combining Lemma 3 and Lemma 4 we have

$$J_{212}=O_p\left(h^2+\frac{1}{\sqrt{nh}}\right)\cdot\{O_p(n^{-1})+O_p(k_n/\sqrt{n}+k_n^{-3/2})\}=o_p(n^{-1/2})$$

and $J_{213}=o_p(n^{-1/2})$. Further,

$$\begin{aligned}J_{211}\ =\ &\tfrac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\mathbf{X}_{\iota(g,i_1)}^T\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)\\ &+\tfrac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))\end{aligned}$$

It is easy to see that

$$E\left\{\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\frac{1}{I_g-1}\sum_{i_1=2}^{I_g}\mathbf{X}_{\iota(g,i_1)}^T\right\}^{\otimes 2}=\sum_{g=1}^{G}\sum_{i=1}^{I_g}E\left(\mathbf{\Pi}_{\iota(g,i)}\frac{1}{I_g-1}\sum_{i_1=2}^{I_g}\mathbf{X}_{\iota(g,i_1)}^T\right)^{\otimes 2}=O(n)$$

where $\mathbf{A}^{\otimes}$ means $\mathbf{A}^T\mathbf{A}$. Combining the root-n consistency of $\tilde{\boldsymbol{\beta}}_n$; it holds that

$$\frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\frac{1}{I_g-1}\sum_{i_1=2}^{I_g}\mathbf{X}_{\iota(g,i_1)}^T\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)=O_p(n^{-1}).$$

According to the definition of $\tilde{m}_n(\cdot)$ we have

$$\begin{aligned}
&\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))\\
=&\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i_1)})-\boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\mathbf{\Xi}^T\mathbf{M_B}\mathbf{\Xi})^{-}\mathbf{\Xi}^T\mathbf{M_B}\mathbf{M})\\
&+\sum_{g=1}^{G}\sum_{i=1}^{I_g}\Pi_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\mathbf{\Xi}^T\mathbf{M_B}\mathbf{\Xi})^{-}\mathbf{\Xi}^T\mathbf{M_B}\mathbf{X}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)\\
&+\sum_{g=1}^{G}\sum_{i=1}^{I_g}\Pi_{\iota(g,i)}\tfrac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\mathbf{\Xi}^T\mathbf{M_B}\mathbf{\Xi})^{-}\mathbf{\Xi}^T\mathbf{M_B}\varepsilon=J_3+J_4+J_5\text{, say.}
\end{aligned}$$

Now, we will prove $J_s = o_p(n^{1/2})$ for $s = 3$, 4 and 5. For convenience, we let

$$\tilde{m}_{\iota(g,i)}=\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i_1)})-\boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\mathbf{\Xi}^T\mathbf{M_B}\mathbf{\Xi})^{-}\mathbf{\Xi}^T\mathbf{M_B}\mathbf{M}).$$

It is easy to see, in order to complete the proof of $J_3 = o_p(n^{1/2})$, we just need to show that $G^{-1}\sum_{g=1}^{G}\Pi_{\iota(g,1),1}\tilde{m}_{\iota(g,1)}=o_p(n^{-1/2})$. Following the proof of Lemma 3 we have

$$\tilde{m}_1=\max_{1\le g\le G}\left|\tilde{m}_{\iota(g,1)}\right|=O\left(\sqrt{k_n}/n+k_n^{-1}\right)\text{ a.s..}$$

Put $\tau_g = \Pi_{\iota(g,1),1}\tilde{m}_{\iota(g,1)}$. For any $\delta > 0$, set

$$\tilde{\Pi}'_{\iota(g,1),1}=\Pi_{\iota(g,1),1}I_{\left\{\left|\Pi_{\iota(g,1),1}\right|\le\delta^2 g^{1/2}\right\}}\quad\text{and}\quad\tilde{\Pi}''_{\iota(g,1),1}=\Pi_{\iota(g,1),1}I_{\left\{\left|\Pi_{\iota(g,1),1}\right|>\delta^2 g^{1/2}\right\}}$$

so that

$$\tau_g=\tilde{m}_{\iota(g,1)}\Pi'_{\iota(g,1),1}+\tilde{m}_{\iota(g,1)}\Pi''_{\iota(g,1),1}.$$

By the three-series theorem we obtain $\sum_{g=1}^{\infty}\left|\Pi''_{\iota(g,1),1}\right|<\infty$ for all $g = 1,\ldots,G$. This implies that

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\Pi''_{\iota(g,1),1}\tilde{m}_{\iota(g,1)}=o\,(1)\ \text{a.s..}$$

For $g = 1,\ldots,G$, let $\tau'_g=\Pi'_{\iota(g,1),1}\tilde{m}_{\iota(g,1)}$. Then given $\tilde{\Delta}_n=\{U_1,\ldots,U_n\}$, $\tau'_1,\ldots,\tau'_G$ are independent and

$$E\left(\tau_g\middle|\tilde{\Delta}_n\right)=0,\ \max_{1\le g\le G}\left|\tau_g\right|\le \tilde{m}_1\delta^2G^{1/2}\quad\text{and}\quad E(\tau_g^2|\tilde{\Delta}_n)=2\tilde{m}_{\iota(g,1)}\sigma^2.$$

By Bernstein's inequality we have

$$\begin{aligned}
p_m &= P\left[\bigcup_{n\ge m}\left\{\frac{1}{\sqrt{G}}\left|\sum_{g=1}^{G}\tau'_g\right|\ge\delta\right\}\right]\le\sum_{G\ge m}E\left[\frac{1}{\sqrt{G}}\sum_{g=1}^{G}Pr\left\{\left|\sum_{g=1}^{G}\tau'_g\right|\ge\delta\middle|\tilde{\Delta}_n\right\}\right]\\
&\le 2\sum_{G\ge m}\sum_{g=1}^{G}E\left[\exp\left\{-\frac{G(\delta/G)^2}{(2/G)\sum_{g=1}^{G}E[(\tau'_g)^2|\tilde{\Delta}_n]+\delta^2G^{1/2}\tilde{m}_1\,(\delta/G)}\right\}\right]\\
&\le 2\sum_{G\ge m}\sum_{g=1}^{G}E\left[\exp\left\{\frac{\delta^2}{2\delta^3G^{1/2}\tilde{m}_1}\right\}\right]\le 2\sum_{G\ge m}G^{-2}\to 0
\end{aligned}$$

as $m\to\infty$. By this we have

$$Pr\left(\left|\frac{1}{\sqrt{G}}\sum_{g=1}^{G}\tau'_g\right|\ge\delta\right)\le p_m+Pr\,(\tilde{m}_1\ge\delta^2)\le 2\delta.$$

Therefore, $J_3 = o(n^{1/2})$ a.s.

By the Cauchy-Schwarz inequality, it holds that

$$\begin{aligned}
\|J_4\| &= \frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\mathbf{\Pi}^T_{\iota(g,i)}\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\mathbf{\Xi}^T\mathbf{M_B}\mathbf{\Xi})^-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\mathbf{\Pi}^T_{\iota(g,i)}\\
&\quad+\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)^T\mathbf{X}^T\mathbf{M_B}\mathbf{\Xi}(\mathbf{\Xi}^T\mathbf{M_B}\mathbf{\Xi})^-\mathbf{\Xi}^T\mathbf{M_B}\mathbf{X}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)=S_1+S_2,\ \text{say.}
\end{aligned}$$

Further,

$$S_1=O(n^{-1}k_n)\cdot\frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\|\mathbf{\Pi}^T_{\iota(g,i)}\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\boldsymbol{\zeta}^T(U_{\iota(g,i_1)})\right\|=O_P(k_n^2)=o_P\,(n)\,.$$

and

$$S_2 \le O_p(1) \cdot \left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n\right)^T \mathbf{X}^T \mathbf{X} \left(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}_n\right) = O_p(1) = o_p(n).$$

Thus, $J_4 = o_p(n^{1/2})$.

In addition, it holds that

$$\begin{aligned} J_5 &= \frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{I_g} \boldsymbol{\Pi}_{\iota(g,i)}^T \frac{1}{I_g-1} \sum_{i_1=1, i_1\neq i}^{I_g} \boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\boldsymbol{\Xi}^T \mathbf{M}_{\mathbf{B}} \boldsymbol{\Xi})^- \frac{1}{I_g-1} \sum_{i_1=1, i\neq 1}^{I_g} \boldsymbol{\zeta}(U_{\iota(g,i_1)}) \boldsymbol{\Pi}_{\iota(g,i)} \\ &\quad + \frac{1}{2}\sum_{g=1}^{G}\sum_{i=1}^{I_g} \boldsymbol{\varepsilon}^T \frac{1}{I_g-1} \sum_{i_1=1, i_1\neq i}^{I_g} \boldsymbol{\zeta}^T(U_{\iota(g,i_1)})(\boldsymbol{\Xi}^T \mathbf{M}_{\mathbf{B}} \boldsymbol{\Xi})^- \frac{1}{I_g-1} \sum_{i_1=1, i\neq 1}^{I_g} \boldsymbol{\zeta}(U_{\iota(g,i_1)}) \boldsymbol{\varepsilon} \\ &= S_3 + S_4, \text{ say.} \end{aligned}$$

It is easy to see that

$$ES_3 \le O(1) \cdot E\left\{\boldsymbol{\Pi} \mathbf{M}_{\mathbf{B}} \boldsymbol{\Xi} (\boldsymbol{\Xi}^T \mathbf{M}_{\mathbf{B}} \boldsymbol{\Xi})^- \boldsymbol{\Xi}^T \mathbf{M}_{\mathbf{B}} \boldsymbol{\Pi}^T\right\} = O(k_n) = o(n^{\frac{1}{2}}).$$

This implies that $S_3 = o_p(n^{1/2})$. Following the same line, we can show that $S_4 = o_p(n^{1/2})$. So $J_5 = o_p(n^{1/2})$ holds. In summary, the proof of Theorem 1 completes.

Proof of Theorem 2

According to the definition of $\hat{m}_n(u)$ it holds that

$$\begin{aligned} \widehat{m}_n(u) &= (1,0)(\mathbf{D}_u^T \mathbf{W}_u \mathbf{D}_u)^{-1} \mathbf{D}_u^T \mathbf{W}_u \\ &\quad \cdot \left\{ \boldsymbol{\varepsilon} - \left(\frac{1}{I_1-1}\sum_{i_1=2}^{I_1} \varepsilon_\iota(1,i_1), \ldots, \frac{1}{I_1-1}\sum_{i_1=1, i_1\neq I_1}^{I_1} \varepsilon_\iota(1,i_1), \ldots, \frac{1}{I_G-1}\sum_{i_1=1, i_1\neq I_G}^{I_G} \varepsilon_\iota(G,i_1) \right)^T \right\} \\ &\quad + (1,0)(\mathbf{D}_u^T \mathbf{W}_u \mathbf{D}_u)^{-1} \mathbf{D}_u^T \mathbf{W}_u (\Delta_1(-1), \ldots, \Delta_1(-I_1), \ldots, \nabla_G(-I_G))^T \\ &\quad + (1,0)(\mathbf{D}_u^T \mathbf{W}_u \mathbf{D}_u)^{-1} \mathbf{D}_u^T \mathbf{W}_u \mathbf{X}(\boldsymbol{\beta} - \widehat{\boldsymbol{\beta}}_n) + (1,0)(\mathbf{D}_u^T \mathbf{W}_u \mathbf{D}_u)^{-1} \mathbf{D}_u^T \mathbf{W}_u \mathbf{M} \\ &= J_1 + J_2 + J_3 + J_4, \text{ say.} \end{aligned}$$

It is easy to see that

$$\mathbf{D}_u^T \mathbf{W}_u \mathbf{D}_u = \begin{pmatrix} \sum_{i=1}^n K_h(U_i - u) & \sum_{i=1}^n \left(\frac{U_i - u}{h}\right) K_h(U_i - u) \\ \sum_{i=1}^n \left(\frac{U_i - u}{h}\right) K_h(U_i - u) & \sum_{i=1}^n \left(\frac{U_i - u}{h}\right)^2 K_h(U_i - u) \end{pmatrix}.$$

Each element of the above matrix is in the form of kernel regression. By Lemma 1 it holds that

$$\mathbf{D}_u^T \mathbf{W}_u \mathbf{D}_u = np(u) \otimes \begin{pmatrix} 1 & 0 \\ 0 & \mu_2 \end{pmatrix} \left[1 + \left\{ \frac{\log(1/h)}{nh} \right\}^{\frac{1}{2}} \right]$$

holds uniformly in u, where ⊗ is the Kronecker product and $\mu_2 = \int_{\mathcal{U}} u^2 K(u) du$ By using the same argument, we have

$$\mathbf{D}_u^T\mathbf{W}_u\mathbf{X}=np\,(u)\,E(\mathbf{1}^T\mathbf{X}_1\big|U)\otimes\begin{pmatrix}1 & 0\\ 0 & \mu_2\end{pmatrix}\left[1+\left\{\frac{\log(1/h)}{nh}\right\}^{\frac{1}{2}}\right]$$

Therefore, combining the fact $\|\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n\|=O_p(n^{1/2})$ we have $J_3 = O_p(n^{1/2})$. Moreover, let

$$\Delta_g^{(1)}(-i)=\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\mathbf{X}_{\iota(g,i_1)}^T\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)\text{ and }\Delta_g^{(2)}(-i)=\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))$$

for $g = 1,\ldots,G$. Then, we have

$$\begin{aligned}J_2 \;=\; & (1,0)\,(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\begin{pmatrix}\sum_{g=1}^{G}\sum_{i=1}^{I_g}K_h\left(U_{\iota(g,i)}-u\right)\Delta_g^{(1)}(-i)\\ \sum_{g=1}^{G}\sum_{i=1}^{I_g}\frac{U_{\iota(g,i)}-u}{h}K_h(U_{\iota(g,i)}-u)\,\Delta_g^{(1)}(-i)\end{pmatrix}\\ & +(1,0)\,(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\begin{pmatrix}\sum_{g=1}^{G}\sum_{i=1}^{I_g}K_h(U_{\iota(g,i)}-u)\Delta_g^{(2)}(-i)\\ \sum_{g=1}^{G}\sum_{i=1}^{I_g}\frac{U_{\iota(g,i)}-u}{h}K_h(U_{\iota(g,i)}-u)\Delta_g^{(2)}(-i)\end{pmatrix}=J_{21}+J_{22},\text{ say.}\end{aligned}$$

By the root-n consistency of $\tilde{\boldsymbol{\beta}}_n$ and the argument as proving J_3 it is easy to see $J_{21} = O_p(n^{-1/2})$. Further,

$$\begin{aligned}& (1,0)\,(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\sum_{g=1}^{G}\sum_{i=1}^{I_g}K_h(U_{\iota(g,i)}-u)\Delta_g^{(2)}(-i)\\ = \; & (1,0)\,(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\sum_{i=1}^{n}K_h\,(U_i-u)\,(m\,(U_i)-\zeta^T\,(U_i)\,(\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi})^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\mathbf{M})\\ & +(1,0)\,(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\sum_{i=1}^{n}K_h\,(U_i-u)\,\zeta^T\,(U_i)\,(\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi})^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\mathbf{X}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)\\ & +(1,0)\,(\mathbf{D}_u^T\mathbf{W}_u\mathbf{D}_u)^{-1}\sum_{i=1}^{n}K_h\,(U_i-u)\,\zeta^T\,(U_i)\,(\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi})^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\varepsilon=J_5+J_6+J_7\text{ say.}\end{aligned}$$

Applying Lemma 1 and the root-n consistency of $\tilde{\boldsymbol{\beta}}_n$ we can show that $J_6 = o_p(n^{-1/2})$. Moreover, by the same argument as proof of Theorem 1 in Horowitz and Mammen (2004) we can show that $J_5=o_p\left\{h^2+1/\sqrt{nh}\right\}$ and $J_7=o_p\left\{h^2+1/\sqrt{nh}\right\}$. Above all we have $J_4=o_p\left\{h^2+1/\sqrt{nh}\right\}$. .

According to the usual nonparametric regression result we have

$$\sqrt{nh}\left[J_4-m\,(u)-\frac{h^2}{2}\frac{\mu_2^2-\mu_1\mu_3}{\mu_2-\mu_1^2}m''\,(u)\right]\rightarrow_p 0\text{ as }n\rightarrow\infty.$$

Therefore, in order to complete the proof we just need to show that

$$\sqrt{nh}J_1\rightarrow_D N\,(0,\zeta\,(u))\quad\text{as }n\rightarrow\infty.$$

Let

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$Q=\frac{1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left[\alpha_0+\alpha_1\left(\frac{U_{\iota(g,i)}-u}{h}\right)\right]K_h\left(U_{\iota(g,i)}-u\right)\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)$$

where $\alpha_0=\mu_2/\left(\mu_2-\mu_1^2\right)$ and $\alpha_1=-\mu_1/\left(\mu_2-\mu_1^2\right)$. It follows that

$$\sqrt{nh}\left[J_1+J_4-m(u)-\frac{h^2}{2}\frac{\mu_2-\mu_1\mu_3}{\mu_2-\mu_1^2}m''(u)\right]=p^{-1}(u)\sqrt{nh}Q+o_p(1).$$

The variance of $\sqrt{nh}Q$ is

$$\begin{aligned}\mathrm{Var}\left(\sqrt{nh}Q\right) = & \frac{h\sigma^2\alpha_0^2}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\frac{I_g}{I_g-1}EK_h^2(U_{\iota(g,i)}-u)\\ & +\frac{h\sigma^2\alpha_1^2}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\frac{I_g}{I_g-1}E\left\{\left(\frac{U_{\iota(g,i)}-u}{h}\right)^2K_h^2(U_{\iota(g,i)}-u)\right\}\\ & +\frac{h\sigma^2\alpha_0\alpha_1}{n}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\frac{I_g}{I_g-1}E\left\{\left(\frac{U_{\iota(g,i)}-u}{h}\right)K_h^2(U_{\iota(g,i)}-u)\right\}\\ & -\frac{h\sigma^2\alpha_0^2}{n}\sum_{g=1}^{G}\sum_{i_1=1}^{I_g}\sum_{i_2=1}^{I_g}\frac{I_g}{\left(I_g-1\right)^2}E\left\{K_h(U_{\iota(g,i_1)}-u)K_h(U_{\iota(g,i_2)}-u)\right\}\\ & +\frac{h\sigma^2\alpha_1^2}{n}\sum_{g=1}^{G}\sum_{i_1=1}^{I_g}\sum_{i_2=1}^{I_g}\frac{I_g}{\left(I_g-1\right)^2}\\ & \cdot E\left\{\left(\frac{U_{\iota(g,i_1)}-u}{h}\right)K_h(U_{\iota(g,i_1)}-u)\left(\frac{U_{\iota(g,i_2)}-u}{h}\right)K_h(U_{\iota(g,i_2)}-u)\right\}\\ & +\frac{2h\sigma^2\alpha_0\alpha_1}{n}\sum_{g=1}^{G}\sum_{i_1=1}^{I_g}\sum_{i_2=1}^{I_g}\frac{I_g}{\left(I_g-1\right)^2}E\left\{\left(\frac{U_{\iota(g,i_1)}-u}{h}\right)K_h^2\left(U_{\iota(g,i_2)}-u\right)\right\}\\ = & J_8+J_9+J_{10}+J_{11}+J_{12}+J_{13}, \quad \text{say.}\end{aligned}$$

It is easy to see that $J_8\to_p\alpha_0^2\sigma^2\nu_0$, $J_9\to_p\alpha_1^2\sigma^2\nu_2$, $J_{10}\to_p\alpha_1\alpha_0\sigma^2\nu_1$, and $J_s\to 0$ for s = 11, 12, 13 as $n\to\infty$. Above all,

$$\mathrm{Var}\left(\sqrt{nh}Q\right)=p^{-1}(u)(\alpha_0^2\nu_0+2\alpha_0\alpha_1\nu_1+\alpha_1^2\nu_2)+o(1).$$

Let

$$a_g=\sqrt{h}\sum_{i=1}^{I_g}\left[\alpha_0+\alpha_1\left(\frac{U_{\iota(g,i)}-u}{h}\right)\right]K_h(U_{\iota(g,i)}-u)\left(\varepsilon_{\iota(g,i)}-\frac{I_g}{I_g-1}\sum_{i_1=1,i_1\neq 1}^{I_g}\varepsilon_{\iota(g,i)}\right)$$

and $B_n^2=\sum_{i=1}^{G}Ea_g^2$. Then

$$B_n^2=np^{-1}(u)(\alpha_0^2\nu_0+2\alpha_0\alpha_1\nu_1+\alpha_1^2\nu_2)\sigma^2+o(n).$$

Simple calculation show that

$$\sum_{g=1}^{G} E\left|a_g\right|^3 \le O(1)\cdot\sum_{g=1}^{G}\sum_{i=1}^{I_g} h^{\frac{3}{2}}\left[|\alpha_0| + |\alpha_1|\cdot\left|\frac{U_{\iota(g,i)}-u}{h}\right|\right] K_h^3(U_{\iota(g,i)}-u)=O(nh^{-1/2}).$$

It follows that $\lim_{n\to\infty} B_n^{-3}\sum_{i=1}^{G} E\left|a_g^3\right|=0.$ By the central limit theorem the proof is complete.

Proof of Theorem 3

For convenience, let

$$\nabla_{\iota(g,i)}=\mathbf{X}_{\iota(g,i)}^T\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right)+m(U_{\iota(g,i)})-\tilde{m}(U_{\iota(g,i)})$$

and

$$\nabla^*_{\iota(g,i)}=\mathbf{X}_{\iota(g,i)}^T(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)+m\left(U_{\iota(g,i)}\right)-\widehat{m}(U_{\iota(g,i)}).$$

By the definition of $\widehat{\sigma}_n^2$ it can be decomposed as

$$\begin{aligned}
\widehat{\sigma}_n^2 &= d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\varepsilon_{\iota(g,i_1)}\right\}^2+d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{\nabla_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\nabla_{\iota(g,i_1)}\right\}^2\\
&+d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\nabla^*_{\iota(g,i)}+2d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\varepsilon_{\iota(g,i_1)}\right\}\\
&\cdot\left\{\nabla_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne 1}^{I_g}\nabla_{\iota(g,i_1)}\right\}+2d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\varepsilon_{\iota(g,i_1)}\right\}\nabla^*_{\iota(g,i)}\\
&+2d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left\{\nabla_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\ne i}^{I_g}\nabla_{\iota(g,i_1)}\right\}\nabla^*_{\iota(g,i)}\\
&= J_1+\ldots+J_6,\quad \text{say}
\end{aligned}$$

where $d(n)=1/\left\{n+\sum_{g=1}^{G} I_g/\left(I_g-1\right)\right\}.$ Applying Lemma 3 and Lemma 4, and Theorem 1 and Theorem 2 it is easy to show that $J_s = o_p(n^{-1/2})$ for s = 2, 3 and 6.

Let

$$\zeta_g=\sum_{i=1}^{I_g}\left\{\varepsilon_{\iota(g,i)}-\left(I_g-1\right)^{-1}\sum_{i_1=1,i_1\ne i}^{I_g}\varepsilon_{\iota(g,i_1)}\right\}^2.$$

Obviously, ζ_g's are independent random variables with $E\zeta_g=\left(I_g-1\right)^{-1}I_g^2\sigma^2.$ Further,

$$\begin{aligned}
& E\left[\sum_{i=1}^{I_g}\left\{\varepsilon_{\iota(g,i)}^2-\frac{1}{(I_g-1)^2}\left(\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)^2-\frac{2}{I_g-1}\varepsilon_{\iota(g,i)}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right\}\right]^2 \\
= & E\left(\sum_{i=1}^{I_g}\varepsilon_{\iota(g,i)}^2\right)^2+E\left[\sum_{i=1}^{I_g}\frac{1}{(I_g-1)^2}\left(\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)^2\right]^2 \\
& +E\left[\frac{2}{I_g-1}\sum_{i=1}^{I_g}\varepsilon_{\iota(g,i)}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right]^2+E\left[\frac{2}{(I_g-1)^2}\sum_{i_1=1}^{I_g}\sum_{i_3=1}^{I_g}\varepsilon_{\iota(g,i_1)}^2\left(\sum_{i_4=1,i_4\neq i_3}^{I_g}\varepsilon_{\iota(g,i_4)}\right)^2\right] \\
& -E\left[\frac{4}{I_g-1}\sum_{i_1=1}^{I_g}\sum_{i_3=1}^{I_g}\varepsilon_{\iota(g,i_1)}^2\varepsilon_{\iota(g,i_3)}\sum_{i_4=1,i_4\neq i_3}^{I_g}\varepsilon_{\iota(g,i_4)}\right] \\
& -E\left[\frac{4}{(I_g-1)^3}\sum_{i_1=1}^{I_g}\sum_{i_3=1}^{I_g}\left(\sum_{i_2=1,i_2\neq i_1}^{I_g}\varepsilon_{\iota(g,i_2)}\right)^2\varepsilon_{\iota(g,i_3)}\sum_{i_4=1,i_4\neq i_3}^{I_g}\varepsilon_{\iota(g,i_4)}\right]=J_7+\ldots+J_{12},\quad \text{say.}
\end{aligned}$$

It is easy to see that

$$\begin{gathered}
J_7=I_gE(\varepsilon_1^4)+I_g\left(I_g-1\right)\sigma^4, \\
J_8=\left(I_g-1\right)^{-2}I_gE(\varepsilon_1^4)+2I_g\left(I_g-1\right)^{-3}[3\left(I_g-2\right)^2+\left(I_g-1\right)]\sigma^4, \\
J_9=12I_g\left(I_g-1\right)^{-1}\sigma^4,\ J_{10}=2\left(I_g-1\right)^{-1}I_gE(\varepsilon_1^4)+2I_g\sigma^4, \\
J_{11}=0,\quad \text{and}\quad J_{12}=-16I_g\left(I_g-1\right)^{-2}\left(I_g-2\right)\sigma^4.
\end{gathered}$$

In summary, we have

$$\begin{aligned}
E(\zeta_g^2) = & \left\{1+\left(I_g-1\right)^{-2}+2\left(I_g-1\right)^{-1}\right\}I_gE(\varepsilon_1^4) \\
& +\left(I_g-1\right)^{-3}(I_g^5-2I_g^4+2I_g^3+6I_g^2+I_g)\sigma^4.
\end{aligned}$$

Then, by some simple calculation, we have

$$\begin{aligned}
\mathrm{Var}(\zeta_g) = & E(\zeta_g^2)-\left\{E(\zeta_g)\right\}^2=\left\{1+\left(I_g-1\right)^{-2}+2\left(I_g-1\right)^{-1}\right\}I_gE(\varepsilon_1^4) \\
& +\left(I_g-1\right)^{-3}(-I_g^4+2I_g^3+6I_g^2+I_g)\sigma^4.
\end{aligned}$$

Therefore,

$$\begin{aligned}
& \mathrm{Var}\left(\frac{\sqrt{n}}{n+\sum_{g=1}^{G}I_g/\left(I_g-1\right)}\sum_{i=1}^{G}\zeta_g\right)=\frac{n}{\{n+\sum_{g=1}^{G}I_g/\left(I_g-1\right)\}^2} \\
& \cdot\sum_{i=1}^{G}\left[\left\{1+\frac{1}{\left(I_g-1\right)^2}+\frac{2}{\left(I_g-1\right)}\right\}I_gE(\varepsilon_1^4)+\frac{1}{\left(I_g-1\right)^3}(-I_g^4+2I_g^3+6I_g^2+I_g)\sigma^4\right].
\end{aligned}$$

According to the definition, J_4 can be written as

$$\begin{aligned} J_4 &= d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\mathbf{X}_{\iota(g,i)}^{T}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right) \\ &\quad -d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\mathbf{X}_{\iota(g,i_1)}^{T}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right) \\ &\quad +d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)(m(U_{\iota(g,i)})-\tilde{m}(U_{\iota(g,i)})) \\ &\quad -d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq 1}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i)})-\tilde{m}(U_{\iota(g,i)})) \\ &= J_{41}-J_{42}+J_{43}-J_{44}. \quad \text{say} \end{aligned}$$

By the proof of Theorem 1, we can show that

$$d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\mathbf{X}_{\iota(g,i)}^{T}=O_p(n^{-\frac{1}{2}}).$$

Therefore, combining the root-n consistency of $\hat{\boldsymbol{\beta}}_n$ we have $J_{41} = o_p(n^{-1/2})$. By the same argument we can show that $J_{42} = o_p(n^{-1/2})$. Further, it holds that

$$\begin{aligned} &J_{43} \\ &= d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\left\{m(U_{\iota(g,i)})-\zeta(U_{\iota(g,i)})^T\left\{\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}\right\}^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\mathbf{M}\right\} \\ &\quad +d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\zeta(U_{\iota(g,i)})^T\left\{\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}\right\}^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\mathbf{X}\left(\boldsymbol{\beta}-\tilde{\boldsymbol{\beta}}_n\right) \\ &\quad +d(n)\sum_{g=1}^{G}\sum_{i=1}^{I_g}\left(\varepsilon_{\iota(g,i)}-\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}\varepsilon_{\iota(g,i_1)}\right)\zeta(U_{\iota(g,i)})^T\left\{\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\boldsymbol{\Xi}\right\}^{-}\boldsymbol{\Xi}^T\mathbf{M}_{\mathbf{B}}\varepsilon \\ &= J_{421}+J_{422}+J_{423} \quad \text{say.} \end{aligned}$$

Following the same line as proving

$$\sum_{g=1}^{G}\sum_{i=1}^{I_g}\boldsymbol{\Pi}_{\iota(g,i)}\frac{1}{I_g-1}\sum_{i_1=1,i_1\neq i}^{I_g}(m(U_{\iota(g,i_1)})-\tilde{m}_n(U_{\iota(g,i_1)}))=o_p(n^{1/2})$$

in the proof of Theorem 1, we have $J_{42s} = o_p(n^{-1/2})$ for $s = 1$, 2 and 3. Thus $J_4 = o_p(n^{-1/2})$. By the same argument, we can show that $J_5 = o_p(n^{-1/2})$. The proof of theorem completes.

Proof of Theorem 4

Proving the consistency of $\hat{\sum}$ is trivial. We here omit the detail. We just show the second result. To facilitate the notation we write

$$\nabla_{\iota(g,i)}=(\mathbf{X}_{\iota(g,i)}^{T}-\mathbf{X}_{\iota(g,i-1)}^{T})(\boldsymbol{\beta}-\widehat{\boldsymbol{\beta}}_n)+(m(U_{\iota(g,i)})-m(U_{\iota(g,i-1)}))-(\widehat{m}_n(U_{\iota(g,i)})-\widehat{m}_n(U_{\iota(g,i-1)})).$$

Then it holds that

NIH-PA Author Manuscript

NIH-PA Author Manuscript

$$\sum_{g=1}^{G}\sum_{i=2}^{I_g}\widehat{\psi}_{\iota(g,i)}^{4}=\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})^4+\sum_{g=1}^{G}\sum_{i=2}^{I_g}\nabla_{\iota(g,i)}^{4}$$
$$+4\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})^3\nabla_{\iota(g,i)}+4\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})\nabla_{\iota(g,i)}^{3}$$
$$+6\sum_{g=1}^{G}\sum_{i=2}^{I_g}\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})^2\nabla_{\iota(g,i)}^{2}=J_1+\ldots+J_5,\quad \text{say}$$

For J_1 we have

$$\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})^4=\sum_{g=1}^{G}\sum_{i=2}^{I_g}\left\{\varepsilon_{\iota(g,i)}^{4}-\varepsilon_{\iota(g,i-1)}^{4}+4\varepsilon_{\iota(g,i)}^{2}\varepsilon_{\iota(g,i-1)}^{2}+2\varepsilon_{\iota(g,i)}^{2}\varepsilon_{\iota(g,i-1)}^{2}\right.$$
$$\left.+4\varepsilon_{\iota(g,i)}^{2}\varepsilon_{\iota(g,i)}\varepsilon_{\iota(g,i-1)}+4\varepsilon_{\iota(g,i-1)}^{2}\varepsilon_{\iota(g,i)}\ \varepsilon_{\iota(g,i-1)}\right\}$$
$$=\sum_{g=1}^{G}[\left(4I_g-2\right)E\varepsilon_1^4+\left\{\left(I_g-1\right)\left(I_g+2\right)+4I_g\right\}\sigma^4]+o_p(n).$$

Combining Theorem 1 and Theorem 2 it is easy to show that $\sum_{g=1}^{G}\sum_{i=1}^{I_g}\nabla_{\iota(g,i)}^{4}=o_p(1)$. Next, according to the Hölder inequality, for s = 1, 2 and 3 we have

$$\left|\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})^s\nabla_{\iota(g,i)}^{4-s}\right|\le\left(\sum_{g=1}^{G}\sum_{i=2}^{I_g}\nabla_{\iota(g,i)}^{s}\right)^{(4-s)/4}\left(\sum_{g=1}^{G}\sum_{i=2}^{I_g}(\varepsilon_{\iota(g,i)}-\varepsilon_{\iota(g,i-1)})^4\right)^{s/4}.$$

Therefore, we can show that $J_i = o_p(n)$ for i = 3,…,5. Thus, the proof is complete.

Proof of Theorems 5 and 6

Following the proof of Theorem 1, we can show that

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{jn}^{(1)A}-\boldsymbol{\beta}_j)=\sum\nolimits_j^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\boldsymbol{\Pi}_{\iota(g,i),j}$$
$$\cdot\left[\varepsilon_{\iota(g,i),j}-\tfrac{1}{(I_gJ-1)}\left\{\sum_{i_1=1,i_1\ne i}^{I_g}\varepsilon_{\iota(g,i_1),j}+\sum_{j_1=1,j_1\ne j}^{J}\sum_{i_1=1}^{I_g}\varepsilon_{\iota(g,i_1),j_1}\right\}\right]+o_p(1)$$

and

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_{jn}^{(2)A}-\boldsymbol{\beta}_j)=\sum\nolimits_j^{-1}\frac{1}{\sqrt{n}}\sum_{g=1}^{G}\sum_{i=1}^{I_g}\boldsymbol{\Pi}_{\iota(g,i),j}\left\{\varepsilon_{\iota(g,i),j}-\frac{1}{I_g(J-1)}\sum_{j_1=1,}^{J}\sum_{i_1=1,i_1\ne1}^{I_g}\varepsilon_{\iota(g,i_1),j_1}\right\}+o_p(1).$$

Therefore, combining the central limit theorem and slustky’s theorem we can show that Theorem 5 and Theorem 6 hold.

Proof of Theorem 7

Applying Theorem 5 and Theorem 6, by the same argument as proving Theorem 2 we can show Theorem 7 holds.

References

1. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Broldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Morre T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisen-beuger DD, Armitage JO, Warnke R, Levy R, Wilson W, Grever MR, Byrd JC, Botstein D, Brown PO, Staudt LM. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–511. [PubMed: 10676951]
2. Baltagi, BH. Econometric Analysis of Panel Data. John Wiley & Sons; 1995.
3. Baltagi BH, Li D. Series Estimation of Partially Linear Panel Data Models with Fixed Effects. Annals of Economics and Finance 2002;3:103–116.
4. Bickel PJ, Kwon J. Inference for semiparametric models: some questions and an answer. With comments and a rejoinder by the authors. Statistics Sinica 2001;11:863–960.
5. Brown PO, Botstein D. Exploring the new world of the genome with microarrays. Nature Genetics 1999;21:33–37. [PubMed: 9915498]
6. Dudoit S, Yang YH, Lu P, Lin DM, Peng V, Nagai J, Speed TP. Normalization for cDNA microarray data: a robust composite method addressing single and multiple slide systematic variation. Nucleic Acids Research 2002;30:e15. [PubMed: 11842121]
7. Entorf H. Random walks with drifts: nonsense regression and spurious fixed-effect estimation. Journal of Econometrics 1997;80:287–296.
8. Fan J, Tam P, Vande Woude G, Ren Y. Normalization and analysis of cDNA micro-arrays using within-array replications applied to neuroblastoma cell response to a cytokine. Proceedings of the National Academy of Science 2004:1135–1140.
9. Fan J, Peng H, Huang T. Semilinear high-dimensional model for normalization of microarray data: a theoretical analysis and partial consistency. Journal of American Statistical Association 2005;471:781–798.
10. Fan J, Huang T. Profile likelihood inferences on semiparametric varying-coefficient partially linear models. Bernoulli 2005;11:1031–1057.
11. Golub TR, Slonim DK, Tamayo P, Huard C, Gassenbeek M, Mesirov P, Celler H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: Class discovery and class prediction by gene expression monitoring. Science 1999;286:531–537. [PubMed: 10521349]
12. Hedenfalk I, Duggan D, Chen Y, Radmacher M, Bittner M, Simon R, Meltzer P, Gusterson B, Esteller M, Kallioniemi O, Wilfond B, Borf A, Trent J. Gene-expression profiles in hereditary breast cancer. The New England Journal of Medicine 2001;344:539–548. [PubMed: 11207349]
13. Honoré BE. Orthogonality conditions for Tobit models with fixed effects and lagged dependent variables. Journal of Econometrics 1994;59:35–61.
14. Horowitz JL, Mammen E. Nonparametric estimation of an additive model with a link function. The Annals of Statistics 2004;32:2412–2443.
15. Huang J, Kuo H, Koroleva I, Zhang C, Soares MB. A semi-linear model for normalization and analyis of cDNA microarray data. Tech Report 321, University of Iowa, Department of Statistics. 2003
16. Huang J, Zhang C. Asymptotic analysis of a two-way smiparametric regression model for microarray data. Statistic Sinica. 2003 to appear.
17. Huang J, Wang D, Zhang C. A two-way semilinear model for normalization and analysis of cDNA microarray data. Journal of the American Statistical Association 2005;471:814–829.
18. Kroll TC, Wölfl S. Ranking: a closer look on globalization methods for normalization of gene expression arrays. Nucleic Acids Research 2002;50:e50. [PubMed: 12034851]
19. Lichtenberg FR. Estimation of the internal adjustment cost model using longitudinal establishment data. Review of Economics and Statistics 1988;70:421–430.

20. Mack YP, Silverman BW. Weak and strong uniform consistency of kernel regression estimates. Z. Wahrsch. Verw. Gebiete 1982;61:405–415.
21. Nguyen DV, Wang N, Carroll RJ. Evaluation of missing value estimation for microarray data. Journal of Data Science 2004;2:347–370.
22. Schaid DJ. Case-parents design for gene-environment intercation. Genetic Epidemiology 1999;16:261–273. [PubMed: 10096689]
23. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary cDNA microarray. Science 1995;270:467–470. [PubMed: 7569999]
24. Tseng GC, Oh MK, Rohlin L, Liao JC, Wong WH. Issues in cDNA microarray analysis: quality filtering, channel normalization, models of variations and assessment of gene effects. Nucleic Acids Research 2001;29:2549–2557. [PubMed: 11410663]
25. You J, Zhou Y, Zhou X. Series Estimation in Partially Linear In-slide Regression Models. Journal of the Royal Statistical Society, Ser B. 2005 Submitted to.

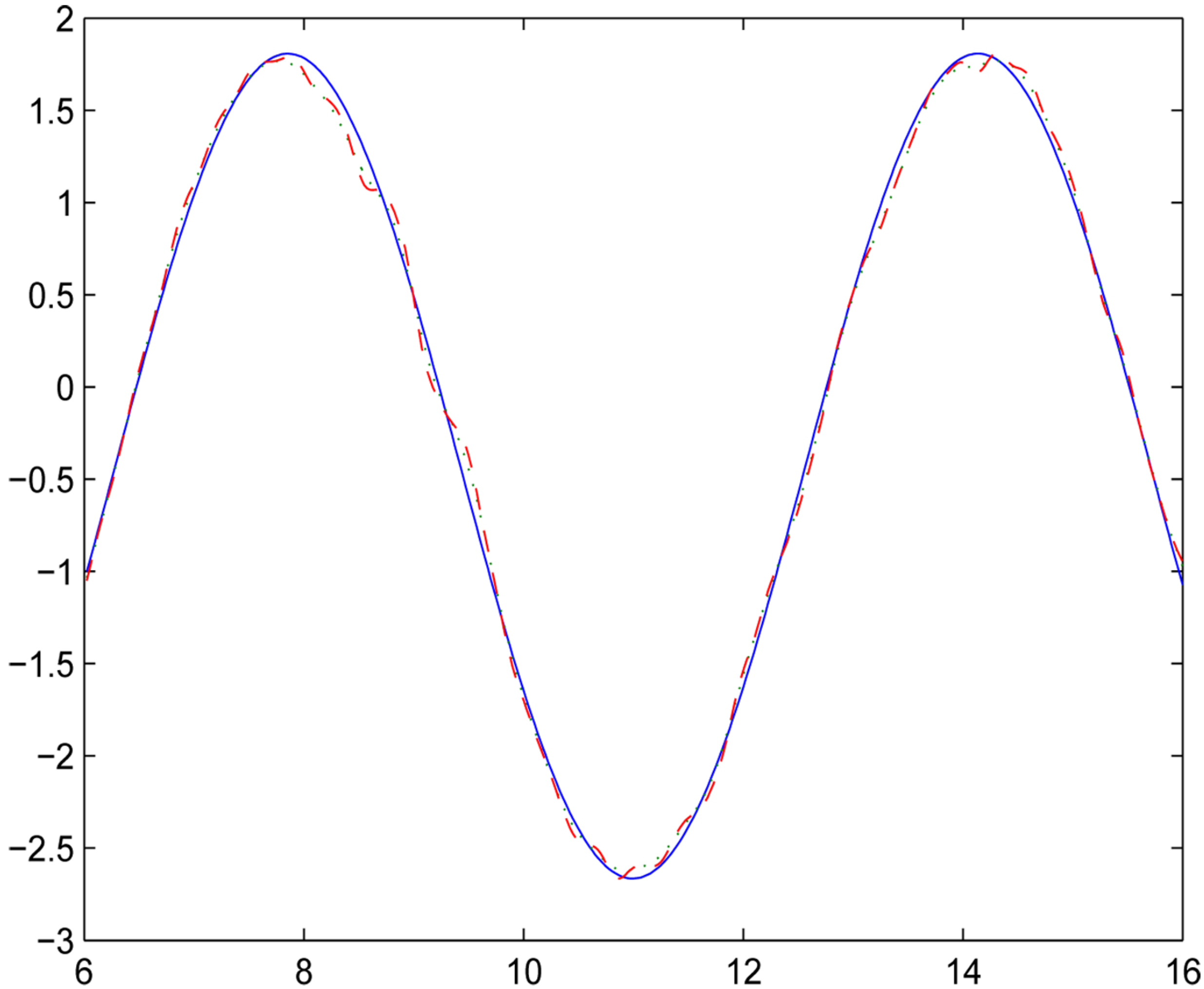

Figure 1.
The estimators of $m(\cdot)$ with $G = 200$ and $I = 4$. Dotted line: the proposed estimator; dash-dotted line: Fan, Peng and Huang (2005)'s estimator; and solid line: $m(\cdot)$.

NIH-PA Author Manuscript

NIH-PA Author Manuscript

Table 1

MSEs of Example 1 (non-aggregation). Fan, Peng and Huang (2005)'s estimation and the proposed estimation

Estimation		I	G=100	G=200	G=400	G=800
Proposed Estimation	$m(\cdot)$	2	0.1451	0.0742	0.0369	0.0208
		3	0.0767	0.0380	0.0233	0.0132
		4	0.0517	0.0269	0.0167	0.0991
	β	2	0.0670	0.0287	0.0156	0.0070
		3	0.0316	0.0149	0.0074	0.0032
		4	0.0214	0.0100	0.0056	0.0020
Fan, Peng and Huang (2005)'s estimation	$m(\cdot)$	2	0.1454	0.0752	0.0358	0.0201
		3	0.0780	0.0397	0.0234	0.0137
		4	0.0515	0.0273	0.0167	0.0100
	β	2	0.0668	0.0299	0.0151	0.0069
		3	0.0318	0.0148	0.0071	0.0033
		4	0.0211	0.0098	0.0050	0.0024