

SFB 649 Discussion Paper 2009-050

Generalized single-index models: The EFM approach

Xia Cui*
Wolfgang Karl Härdle**
Lixing Zhu***



* Sun Yat-Sen University Guangzhou, China
** Humboldt-Universität zu Berlin, Germany
*** Hong Kong Baptist University, Hong Kong

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

<http://sfb649.wiwi.hu-berlin.de>
ISSN 1860-5664

SFB 649, Humboldt-Universität zu Berlin
Spandauer Straße 1, D-10178 Berlin



SFB 649 ECONOMIC RISK BERLIN

Generalized single-index models: The EFM approach *

Xia Cui (cuixia@mail.sysu.edu.cn)

School of Mathematics & Computational Science

Sun Yat-Sen University, Guangzhou, China

Wolfgang Karl Härdle (haerdle@wiwi.hu-berlin.de)

CASE-Center for Applied Statistics & Economics

Institut für Statistik und Ökonometrie

Wirtschaftswissenschaftliche Fakultät

Humboldt-Universität zu Berlin

D-10178 Berlin, Germany

National Central University, Taipei, Taiwan

Lixing Zhu (lzhu@hkbu.edu.hk)

Department of Mathematics, Hong Kong Baptist University, Hong Kong

School of Finance and Statistics, East China Normal University, Shanghai, China

*The authors made equal contributions to this paper. Correspondence should be addressed to Lixing Zhu. The second author thanks the Deutsche Forschungsgemeinschaft SFB 649 “Ökonomisches Risiko” for financial support. The third author thanks the “HKBU 2030/07P” for financial support.

Abstract. Generalized single-index models are natural extensions of linear models and circumvent the so-called curse of dimensionality. They are becoming increasingly popular in many scientific fields including biostatistics, medicine, economics and financial econometrics. Estimating and testing the model index coefficients β is one of the most important objectives in the statistical analysis. However, the commonly used assumption on the index coefficients, $\|\beta\| = 1$, represents a non-regular problem: the true index is on the boundary of the unit ball. In this paper we introduce the EFM approach, a method of estimating functions, to study the generalized single-index model. The procedure is to first relax the equality constraint to one with $(d - 1)$ components of β lying in an open unit ball, and then to construct the associated $(d - 1)$ estimating functions by projecting the score function to the linear space spanned by the residuals with the unknown link being estimated by kernel estimating functions. The root- n consistency and asymptotic normality for the estimator obtained from solving the resulting estimating equations is achieved, and a Wilk's type theorem for testing the index is demonstrated. A noticeable result we obtain is that our estimator for β has smaller or equal limiting variance than the estimator of Carroll et al. (1997). A fixed point iterative scheme for computing this estimator is proposed. This algorithm only involves one-dimensional nonparametric smoothers, thereby avoiding the data sparsity problem caused by high model dimensionality. Numerical studies based on simulation and on applications suggest that this new estimating system is quite powerful and easy to implement.

Key words and phrases: Generalized single-index model, index coefficients, estimating equations, asymptotic properties, iteration.

noindent *JEL-Codes:* C02, C13, C14, C21

1. Introduction

Single-index models combine flexibility of modelling with interpretability of (linear) coefficients. They circumvent the curse of dimensionality and are becoming increasingly popular in many scientific fields. The reduction of dimension is achieved by assuming the link function to be a univariate function applied to the projection of

explanatory covariate vector on to some direction. In this paper we consider an extension of single index models where, instead of a distributional assumption, assumptions of only the mean function and variance function of the response are made. Let $(Y_i, \mathbf{X}_i), i = 1, \dots, n$, denote the observed values with Y_i being the response variable and \mathbf{X}_i as the vector of d explanatory variables. The relationship of the mean and variance of Y_i are specified as follows:

$$E(Y_i|\mathbf{X}_i) = \mu\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}, \text{Var}(Y_i|\mathbf{X}_i) = \sigma^2 V\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}, \quad (1.1)$$

where μ is a known monotonic function, V is a known covariance function, g is an unknown univariate link function and $\boldsymbol{\beta}$ is an unknown index vector which belongs to the parameter space $\Theta = \{\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top : \|\boldsymbol{\beta}\| = 1, \beta_1 > 0, \boldsymbol{\beta} \in \mathbb{R}^d\}$. Here we assume the parameter space is Θ rather than the entire \mathbb{R}^d in order to ensure that $\boldsymbol{\beta}$ in the representation (1.1) can be uniquely defined. This is a commonly used assumption on the index parameter (see Carroll et al., 1997; Zhu and Xue, 2006). Model (1.1) is flexible enough to cover a variety of situations. If μ is the identity function and V is equal to constant 1, (1.1) reduces to a single-index model (Härdle, Hall and Ichimura, 1993). Model (1.1) is an extension of the generalized linear model (McCullagh and Nelder, 1989) and the single index model. When the conditional distribution of Y is logistic, then $\mu\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}/[1 + \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}]$ and $V\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}/[1 + \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}]^2$.

For single-index models: $\mu\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = g(\boldsymbol{\beta}^\top \mathbf{X})$ and $V\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = 1$, various strategies for estimating $\boldsymbol{\beta}$ have been proposed in the last decades. Two most popular methods are the average derivative method (ADE) introduced in Powell et al. (1989) and Härdle and Stoker (1989), and the simultaneous minimization method of Härdle, Hall and Ichimura (1993). Next we will review these two methods in short. The ADE method is based on that $\partial E(Y|\mathbf{X} = \mathbf{x})/\partial \mathbf{x} = g'(\boldsymbol{\beta}^\top \mathbf{x})\boldsymbol{\beta}$ which implies that the gradient of the regression function is proportional to the index parameter $\boldsymbol{\beta}$. Then a natural estimator for $\boldsymbol{\beta}$ is $\hat{\boldsymbol{\beta}} = n^{-1} \sum_{i=1}^n \widehat{\nabla G}(\mathbf{X}_i) / \|n^{-1} \sum_{i=1}^n \widehat{\nabla G}(\mathbf{X}_i)\|$ with $\nabla G(\mathbf{x})$ denoting $\partial E(Y|\mathbf{X} = \mathbf{x})/\partial \mathbf{x}$ and $\|\cdot\|$ being the Euclidean norm. An advantage of the ADE approach is that it allows estimating $\boldsymbol{\beta}$ directly. However, the high-dimensional kernel smoothing used for computing $\widehat{\nabla G}(\mathbf{x})$ suffers from the ‘‘curse of dimensionality’’ if

the model dimension d is large. Hristache, Juditski and Spokoiny (2001) improved the ADE approach by lowering the dimension of the kernel gradually. The method of Härdle, Hall and Ichimura (1993) is carried out by minimizing a least squares criterion based on nonparametric estimation of the link g with respect to $\boldsymbol{\beta}$ and bandwidth h . However, the minimization is difficult to implement since it depends on an optimization problem in a high-dimensional space. Xia et al. (2002) proposed to minimize average conditional variance (MAVE). Because the kernel used for computing $\boldsymbol{\beta}$ is a function of $\|\mathbf{X}_i - \mathbf{X}_j\|$, MAVE meets the problem of data sparseness. All the above estimators are consistent under some regular conditions. Asymptotic efficiency comparisons of the above methods have been discussed in Xia (2006) resulting in the MAVE estimator of $\boldsymbol{\beta}$ having the same limiting variance as the estimators of Härdle, Hall and Ichimura (1993), and claiming that alternative versions of the ADE method having larger variance.

The main challenges of estimation in the semiparametric model (1.1) are that, the support of the infinite dimensional nuisance parameter $g(\cdot)$ depends on the finite dimensional parameter $\boldsymbol{\beta}$, and the parameter $\boldsymbol{\beta}$ is on the boundary of a unit ball. For estimating $\boldsymbol{\beta}$ the former challenge forces us to deal with the infinite dimensional nuisance parameter g . The latter one represents a non-regular problem. The classic assumptions about asymptotic properties of the estimates for $\boldsymbol{\beta}$ are not valid. In addition, as a model proposed for dimension reduction, the dimension d may be very high and one often meets the problem of computation. To attack the above problems, in this paper we will develop an estimating function method (EFM) and then introduce a computational algorithm to solve the equations based on a fixed point iterative scheme. We first choose an identifiable parametrization which transforms the boundary of a unit ball in \mathbb{R}^d to the interior of a unit ball in \mathbb{R}^{d-1} . By eliminating β_1 , the parameter space Θ can be rearranged to a form $\{(1 - \sum_{r=2}^d \beta_r^2)^{1/2}, \beta_2, \dots, \beta_d\}^\top : \sum_{r=2}^d \beta_r^2 < 1\}$. Then the derivatives of a function with respect to $(\beta_2, \dots, \beta_d)^\top$ are readily obtained by chain rule and the classic assumptions on the asymptotic normality hold after transformation. The estimating functions (equations) for $\boldsymbol{\beta}$ can be constructed by replacing $g(\boldsymbol{\beta}^\top \mathbf{X})$ with $\hat{g}(\boldsymbol{\beta}^\top \mathbf{X})$. The estimate \hat{g} for the nuisance parameter g is obtained using kernel estimating functions and the smoothing parameter h is selected using K -fold cross-validation. For the problem of testing the index, we establish a quasi-likelihood ratio

based on the proposed estimating functions and show that the test statistics asymptotically follow a χ^2 -distribution whose degree of freedom does not depend on nuisance parameters, under the null hypothesis. Then a Wilks' type theorem for testing the index is demonstrated.

The proposed EFM technique is essentially a unified method of handling different types of data situations including categorical response variable and discrete explanatory covariate vector. The main results of this research are as follows:

- (a) *Efficiency.* A surprising result we obtain is that our EFM estimator for β has smaller or equal limiting variance than the estimator of Carroll et al. (1997).
- (b) *Computation.* The estimating function system only involves one-dimensional non-parametric smoothers, thereby avoiding the data sparsity problem caused by high model dimensionality. Unlike the quasi-likelihood inference (Carroll et al., 1997) where the maximization is difficult to implement when d is large, the reparametrization and the explicit formulation of the estimating functions facilitate an efficient computation algorithm. Here we use a fixed point iterative scheme to compute the resultant estimator. The simulation results show that the algorithm adapts to higher model dimension and richer data situations than the MAVE method of (Xia et al., 2002).

The paper is organized as follows. In Section 2, we state the generalized single-index model, discuss estimation of g using kernel estimating functions and of β using profile estimating functions, and investigate the problem of testing the index using quasi-likelihood ratio. In Section 3 we provide a computation algorithm for solving the estimating functions and illustrate the method with simulation and practical studies. The proofs are deferred to the Appendix.

2. Estimating function method (EFM) and its large sample properties

In this section, which is concerned with inference based on the estimating function method, the model of interest is determined through specification of mean and variance functions, up to an unknown vector β and an unknown function g . Except for Gaussian

data, model (1.1) needs not be a full semiparametric likelihood specification. Note that the parameter space $\Theta = \{\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)^\top : \|\boldsymbol{\beta}\| = 1, \beta_1 > 0, \boldsymbol{\beta} \in \mathbb{R}^d\}$ means that $\boldsymbol{\beta}$ is on the boundary of a unit ball and it represents therefore a non-regular problem. So we first choose an identifiable parametrization which transforms the boundary of a unit ball in \mathbb{R}^d to the interior of a unit ball in \mathbb{R}^{d-1} . By eliminating β_1 , the parameter space Θ can be rearranged to a form $\{((1 - \sum_{r=2}^d \beta_r^2)^{1/2}, \beta_2, \dots, \beta_d)^\top : \sum_{r=2}^d \beta_r^2 < 1\}$. Then the derivatives of a function with respect to $\boldsymbol{\beta}^{(1)} = (\beta_2, \dots, \beta_d)^\top$ are readily obtained by chain rule and the classic assumptions on the asymptotic normality hold after transformation. This reparametrization is the key to analyzing the asymptotic properties of the estimates for $\boldsymbol{\beta}$ and to facilitating an efficient computation algorithm. We will investigate the estimation for g and $\boldsymbol{\beta}$ and propose a quasi-likelihood method to test the statistical significance of certain variables in the parametric component.

2.1. The kernel estimating functions for the nonparametric part g

If $\boldsymbol{\beta}$ is known, then we estimate $g(\cdot)$ and $g'(\cdot)$ using the local linear estimating functions. Let h denote the bandwidth parameter, and let $K(\cdot)$ denote the symmetric kernel density function satisfying $K_h(\cdot) = h^{-1}K(\cdot/h)$. The estimation method involves local linear approximation. Denote by α_0 and α_1 the values of g and g' evaluating at t , respectively. The local linear approximation for $g(\boldsymbol{\beta}^\top \mathbf{x})$ in a neighborhood of t is $\tilde{g}(\boldsymbol{\beta}^\top \mathbf{x}) = \alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{x} - t)$. The estimators $\hat{g}(t)$ and $\hat{g}'(t)$ are obtained by solving the kernel estimating functions with respect to α_0, α_1 :

$$\begin{cases} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - t) \mu' \{\tilde{g}(\boldsymbol{\beta}^\top \mathbf{X}_j)\} V^{-1} \{\tilde{g}(\boldsymbol{\beta}^\top \mathbf{X}_j)\} [Y_j - \mu \{\tilde{g}(\boldsymbol{\beta}^\top \mathbf{X}_j)\}] = 0, \\ \sum_{j=1}^n (\boldsymbol{\beta}^\top \mathbf{X}_j - t) K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - t) \mu' \{\tilde{g}(\boldsymbol{\beta}^\top \mathbf{X}_j)\} V^{-1} \{\tilde{g}(\boldsymbol{\beta}^\top \mathbf{X}_j)\} [Y_j - \mu \{\tilde{g}(\boldsymbol{\beta}^\top \mathbf{X}_j)\}] = 0, \end{cases} \quad (2.1)$$

Having estimated α_0, α_1 at t as $\hat{\alpha}_0, \hat{\alpha}_1$, the local linear estimators of $g(t)$ and $g'(t)$ are $\hat{g}(t) = \hat{\alpha}_0$ and $\hat{g}'(t) = \hat{\alpha}_1$ respectively.

The key to obtain the asymptotic normality of the estimates for $\boldsymbol{\beta}$ lies in the asymptotic properties of the estimated nonparametric part. The following theorem will provide some useful results. The following notation will be used. Let $\mathcal{X} = \{\mathbf{X}_1, \dots, \mathbf{X}_n\}$,

$\rho_l(z) = \{\mu'(z)\}^l V^{-1}(z)$ and $\mathbf{J} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{(1)}}$ the Jacobian matrix of size $d \times (d-1)$ with

$$\mathbf{J} = \begin{pmatrix} -\boldsymbol{\beta}^{(1)\top} / \sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2} \\ \mathbf{I}_{d-1} \end{pmatrix}, \quad \boldsymbol{\beta}^{(1)} = (\beta_2, \dots, \beta_d)^\top.$$

The moments of K and K^2 are denoted respectively by, $j = 0, 1, \dots$

$$\gamma_j = \int t^j K(t) dt \quad \text{and} \quad \nu_j = \int t^j K^2(t) dt.$$

Proposition 1. *Under regularity conditions (a), (b), (d) and (e) given in the Appendix, we have*

(i) *With $h \rightarrow 0$, $n \rightarrow \infty$ such that $h \rightarrow 0$ and $nh \rightarrow \infty$, $\forall \boldsymbol{\beta} \in \Theta$, the asymptotic conditional bias and variance of \hat{g} is given by*

$$\begin{aligned} E\left\{\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x})\}^2 \middle| \mathcal{X}\right\} &= \left\{\frac{1}{2}\gamma_2 h^2 g''(\boldsymbol{\beta}^\top \mathbf{x})\right\}^2 \\ &+ \nu_0 \sigma^2 / [nh f_{\boldsymbol{\beta}^\top \mathbf{x}}(\boldsymbol{\beta}^\top \mathbf{x}) \rho_2\{g(\boldsymbol{\beta}^\top \mathbf{x})\}] \quad (2.2) \\ &+ \mathcal{O}_P(h^4 + n^{-1}h^{-1}). \end{aligned}$$

(ii) *With $h \rightarrow 0$, $n \rightarrow \infty$ such that $h \rightarrow 0$ and $nh^3 \rightarrow \infty$, for the estimates of the derivative g' , it holds that*

$$\begin{aligned} E\left\{\{\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x}) - g'(\boldsymbol{\beta}^\top \mathbf{x})\}^2 \middle| \mathcal{X}\right\} &= \left\{\frac{1}{6}\gamma_4 \gamma_2^{-1} h^2 g'''(\boldsymbol{\beta}^\top \mathbf{x}) + \frac{1}{2}(\gamma_4 \gamma_2^{-1} - \gamma_2) h^2 g''(\boldsymbol{\beta}^\top \mathbf{x})\right. \\ &\quad \times [\rho_2'\{g(\boldsymbol{\beta}^\top \mathbf{x})\} / \rho_2\{g(\boldsymbol{\beta}^\top \mathbf{x})\} \\ &\quad \left. + f'_{\boldsymbol{\beta}^\top \mathbf{x}}(\boldsymbol{\beta}^\top \mathbf{x}) / f_{\boldsymbol{\beta}^\top \mathbf{x}}(\boldsymbol{\beta}^\top \mathbf{x})]\right\}^2 \\ &+ \nu_2 \gamma_2^{-2} \sigma^2 / [nh^3 f_{\boldsymbol{\beta}^\top \mathbf{x}}(\boldsymbol{\beta}^\top \mathbf{x}) \rho_2\{g(\boldsymbol{\beta}^\top \mathbf{x})\}] \\ &+ \mathcal{O}_P(h^4 + n^{-1}h^{-3}). \quad (2.3) \end{aligned}$$

(iii) *With $h \rightarrow 0$, $n \rightarrow \infty$ such that $h \rightarrow 0$ and $nh^3 \rightarrow \infty$, we have that*

$$E\left\{\left\|\frac{\partial \hat{g}(\boldsymbol{\beta}^\top \mathbf{x})}{\partial \boldsymbol{\beta}^{(1)}} - g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{\mathbf{x} - E(\mathbf{x} | \boldsymbol{\beta}^\top \mathbf{x})\}\right\|^2 \middle| \mathcal{X}\right\} = \mathcal{O}_P(h^4 + n^{-1}h^{-3}), \quad (2.4)$$

The proof of this Proposition appears in the Appendix. Results (i) and (ii) in Proposition 1 is routine and in principal similar to Carroll, Ruppert and Welsh (1998). In the situation where $\sigma^2 V = \sigma^2$ and the function μ is identity, results (i) and (ii) coincides with that given by Fan and Gijbels (1996). From result (iii), it is seen that

$\partial\hat{g}(\boldsymbol{\beta}^\top \mathbf{x})/\partial\boldsymbol{\beta}^{(1)}$ converges in probability to $g'(\boldsymbol{\beta}^\top \mathbf{x})\mathbf{J}^\top\{\mathbf{x} - E(\mathbf{x}|\boldsymbol{\beta}^\top \mathbf{x})\}$, rather than $g'(\boldsymbol{\beta}^\top \mathbf{x})\mathbf{J}^\top \mathbf{x}$ as if g were known. That is, $\lim_{n \rightarrow \infty} \{\partial\hat{g}(\boldsymbol{\beta}^\top \mathbf{x})/\partial\boldsymbol{\beta}^{(1)}\} \neq \partial\{\lim_{n \rightarrow \infty} \hat{g}(\boldsymbol{\beta}^\top \mathbf{x})\}/\partial\boldsymbol{\beta}^{(1)}$, which means that the convergence in probability and the derivation of the sequence $\hat{g}_n(\boldsymbol{\beta}^\top \mathbf{x})$ (as a function of n) can not commute. This is primarily caused by the fact that the support of the infinite dimensional nuisance parameter $g(\cdot)$ depends on the finite dimensional projection parameter $\boldsymbol{\beta}$. In contrast, a semiparametric model where the support of the nuisance parameter is independent of the finite dimensional parameter is a partially linear regression model having form $Y = \mathbf{X}^\top \boldsymbol{\theta} + \eta(T) + \varepsilon$. It is easy to check that the limit of $\partial\hat{\eta}(T)/\partial\boldsymbol{\theta}$ is equal to $E(\mathbf{X}|T)$, which is the derivative of $\lim_{n \rightarrow \infty} \hat{\eta}(T) = E(Y|T) - E(\mathbf{X}^\top|T)\boldsymbol{\theta}$ with respect to $\boldsymbol{\theta}$. Result (iii) ensures that the proposed estimator does not require undersmoothing of $g(\cdot)$ to obtain a root- n consistent estimator for $\boldsymbol{\beta}$ and it is also of its own interest in inference theory for semiparametric models.

2.2. The asymptotic distribution for the estimates of the parametric part $\boldsymbol{\beta}$

We will now proceed to the estimation of $\boldsymbol{\beta} \in \Theta$. We need to estimate the $(d-1)$ -dimensional vector $\boldsymbol{\beta}^{(1)}$, the estimator of which will be defined via:

$$\sum_{i=1}^n [\partial\mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}/\partial\boldsymbol{\beta}^{(1)}] V^{-1} \{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] = 0. \quad (2.5)$$

This is the direct analogue of the ‘‘ideal’’ estimating equation for known g , in that it is calculated by replacing $g(t)$ with $\hat{g}(t)$. An asymptotically equivalent and easily computed version of this equation is:

$$\hat{\mathbf{G}}(\boldsymbol{\beta}) \stackrel{\text{def}}{=} \sum_{i=1}^n \mathbf{J}^\top \hat{g}'(\boldsymbol{\beta}^\top \mathbf{X}_i) \{\mathbf{X}_i - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \rho_1 \{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}], \quad (2.6)$$

with $\mathbf{J} = \frac{\partial\boldsymbol{\beta}}{\partial\boldsymbol{\beta}^{(1)}}$ the Jacobian mentioned above, \hat{g} and \hat{g}' are defined by (2.1), and $\hat{\mathbf{h}}(t)$ the local linear estimate for $\mathbf{h}(t) = E(\mathbf{X}|\boldsymbol{\beta}^\top \mathbf{X} = t) = (h_1(t), \dots, h_d(t))^\top$,

$$\hat{\mathbf{h}}(t) = \sum_{i=1}^n b_i(t) \mathbf{X}_i / \sum_{i=1}^n b_i(t),$$

where $b_i(t) = K_h(\boldsymbol{\beta}^\top \mathbf{X}_i - t) \{S_{n,2}(t) - (\boldsymbol{\beta}^\top \mathbf{X}_i - t)S_{n,1}(t)\}$ and $S_{n,k} = \sum_{i=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_i - t)(\boldsymbol{\beta}^\top \mathbf{X}_i - t)^k, k = 1, 2$. We use (2.6) to estimate $\boldsymbol{\beta}^{(1)}$ in the generalized single-index

model, and then use the fact that $\beta_1 = \sqrt{1 - \|\boldsymbol{\beta}^{(1)}\|^2}$ to obtain $\hat{\beta}_1$. The use of (2.6) constitutes in our view a new approach to estimating generalized single index models, since (2.6) involves smooth pilot estimation of g , g' and \mathbf{h} we call it the Estimation Function Method (EFM) for $\boldsymbol{\beta}$.

Remark 1. *It can be seen from the proof in the Appendix that the population version of $\hat{\mathbf{G}}(\boldsymbol{\beta})$ is*

$$\mathbf{G}(\boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{J}^\top g'(\boldsymbol{\beta}^\top \mathbf{X}_i) \{\mathbf{X}_i - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \rho_1\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}], \quad (2.7)$$

which is obtained by replacing $\hat{g}, \hat{g}', \hat{\mathbf{h}}$ with g, g', \mathbf{h} in (2.6). One important property of (2.7) is that the second Bartlett identity holds, for any $\boldsymbol{\beta}$:

$$E\{\mathbf{G}(\boldsymbol{\beta})\mathbf{G}^\top(\boldsymbol{\beta})\} = -E\left\{\frac{\partial \mathbf{G}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}}\right\}.$$

This property makes the semiparametric efficiency of the EFM (2.6) possible.

We will focus now on the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}^{(1)}$ derived from (2.6). The reason is that the asymptotic consistency has been achieved for many existing estimators (Härdle, Hall and Ichimura, 1993; Härdle and Stoker, 1989) in the simple single-index model. Let $\boldsymbol{\beta}^0 = (\beta_1^0, \boldsymbol{\beta}^{(1)0^\top})^\top$ denote the true parameter and \mathbf{B}^+ denote the Moore-Penrose inverse of any given matrix \mathbf{B} . We assume in the following context that $\hat{\boldsymbol{\beta}}^{(1)}$ is in a root- n neighborhood of the true parameter $\boldsymbol{\beta}^{(1)0}$, $\hat{\boldsymbol{\beta}}^{(1)} \in \{\boldsymbol{\beta}^{(1)} : \|\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}^{(1)0}\| \leq Cn^{-1/2}\}$ with C is some constant. We have the following asymptotic result for the estimator $\hat{\boldsymbol{\beta}}^{(1)}$.

Theorem 1. *Let $\hat{\boldsymbol{\beta}}^{(1)}$ denote the solution of the estimating function (2.6). With $h \rightarrow 0$, $n \rightarrow \infty$ such that $nh^6 \rightarrow 0$ and $nh^4 \rightarrow \infty$ and regularity conditions given in the Appendix, we have:*

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(1)0}) \xrightarrow{\mathcal{L}} N_{d-1}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}^{(1)0}}), \quad (2.8)$$

where $\boldsymbol{\Sigma}_{\boldsymbol{\beta}^{(1)0}} = \{\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J}\}^+ |_{\boldsymbol{\beta}^{(1)} = \boldsymbol{\beta}^{(1)0}}$, $\mathbf{J} = \frac{\partial \boldsymbol{\beta}}{\partial \boldsymbol{\beta}^{(1)}}$ and

$$\boldsymbol{\Omega} = E\left[\{\mathbf{X}\mathbf{X}^\top - E(\mathbf{X}|\boldsymbol{\beta}^\top \mathbf{X})E(\mathbf{X}^\top|\boldsymbol{\beta}^\top \mathbf{X})\} \rho_2\{g(\boldsymbol{\beta}^\top \mathbf{X})\} \{g'(\boldsymbol{\beta}^\top \mathbf{X})\}^2 / \sigma^2\right].$$

Remark 2. Note that $\boldsymbol{\beta}^\top \boldsymbol{\Omega} \boldsymbol{\beta} = 0$, so the nonnegative matrix $\boldsymbol{\Omega}$ degenerates in the direction of $\boldsymbol{\beta}$. If the mean function μ is the identity function and the variance function is equal to a scale constant, that is, $\mu\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = g(\boldsymbol{\beta}^\top \mathbf{X})$, $\sigma^2 V\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = \sigma^2$, the matrix $\boldsymbol{\Omega}$ in Theorem 1 reduces to be

$$\boldsymbol{\Omega} = E\left[\{\mathbf{X}\mathbf{X}^\top - E(\mathbf{X}|\boldsymbol{\beta}^\top \mathbf{X})E(\mathbf{X}^\top|\boldsymbol{\beta}^\top \mathbf{X})\}\{g'(\boldsymbol{\beta}^\top \mathbf{X})\}^2/\sigma^2\right].$$

Technically speaking, Theorem 1 shows that an undersmoothing approach is unnecessary and that root- n consistency can be achieved. The asymptotic covariance $\boldsymbol{\Sigma}_{\boldsymbol{\beta}(1)0}$ in general can be estimated by replacing terms in its expression by estimates of those terms. The asymptotic normality of $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\boldsymbol{\beta}}^{(1)\top})^\top$ will follow from Theorem 1 with a simple application of the multivariate delta-method, since $\hat{\beta}_1 = \sqrt{1 - \|\hat{\boldsymbol{\beta}}^{(1)}\|^2}$. According to the results of Carroll et al. (1997), the asymptotic variance of their estimator is $\boldsymbol{\Omega}^+$. Define the block partition of matrix $\boldsymbol{\Omega}$ as follows:

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}, \quad (2.9)$$

where $\boldsymbol{\Omega}_{11}$ is a positive constant, $\boldsymbol{\Omega}_{12}$ is a $(d-1)$ -dimensional row vector, $\boldsymbol{\Omega}_{21}$ is a $(d-1)$ -dimensional column vector and $\boldsymbol{\Omega}_{22}$ is a $(d-1) \times (d-1)$ nonnegative definite matrix.

Corollary 1. Under the conditions of Theorem 1, we have

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) \xrightarrow{\mathcal{L}} N_p(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\beta}^0}), \quad (2.10)$$

with $\boldsymbol{\Sigma}_{\boldsymbol{\beta}^0} = \mathbf{J}\{\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J}\}^+ \mathbf{J}^\top \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^0}$. Further,

$$\boldsymbol{\Sigma}_{\boldsymbol{\beta}^0} \leq \boldsymbol{\Omega}^+ \Big|_{\boldsymbol{\beta}=\boldsymbol{\beta}^0},$$

and the stick less-than sign holds when $\det(\boldsymbol{\Omega}_{22}) = 0$. That is, in this case the EFM is more efficient than that of Carroll et al. (1997).

2.3. Profile quasi-likelihood ratio test

In applications, it is important to test the statistical significance of added predictors in a regression model. Here we establish a quasi-likelihood ratio statistic to test

the significance of certain variables in the linear index. The null hypothesis that the model is correct is tested against a full model alternative. Fan and Jiang (2007) gave a recent review about generalized likelihood ratio tests. Bootstrap tests for nonparametric regression, generalized partially linear models and single index models can be found in Härdle and Mammen (1993), Härdle, Mammen and Müller (1998) and Härdle, Mammen and Proenca (2001). Consider the testing problem:

$$H_0 : g(\cdot) = g\left(\sum_{k=1}^r \beta_k X_k\right) \longleftrightarrow H_1 : g(\cdot) = g\left(\sum_{k=1}^r \beta_k X_k + \sum_{k=r+1}^d \beta_k X_k\right). \quad (2.11)$$

We mainly focus on testing $\beta_k = 0, k = r+1, \dots, d$, though the following test procedure can be easily extended to a general linear testing $\mathbf{B}\tilde{\boldsymbol{\beta}} = 0$ where \mathbf{B} is a known matrix with full row rank and $\tilde{\boldsymbol{\beta}} = (\beta_{r+1}, \dots, \beta_d)^\top$. The profile quasi-likelihood ratio test is defined by

$$T_n = 2\left\{\sup_{\boldsymbol{\beta} \in \Theta} \hat{Q}(\boldsymbol{\beta}) - \sup_{\boldsymbol{\beta} \in \Theta, \tilde{\boldsymbol{\beta}}=0} \hat{Q}(\boldsymbol{\beta})\right\}, \quad (2.12)$$

where $\hat{Q}(\boldsymbol{\beta}) = \sum_{i=1}^n Q[\mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}, Y_i]$, $Q[\mu, y] = \int_{\mu}^y V\{\mu^{-1}(s)\}(s-y)ds$ and $\mu^{-1}(\cdot)$ is the inverse function of $\mu(\cdot)$. The following Wilks' type theorem shows that the distribution of T_n is asymptotically chi-squared and independent of nuisance parameters.

Theorem 2. *Under the assumptions of Theorem 1, if $\beta_k = 0, k = r+1, \dots, d$, then*

$$T_n \xrightarrow{\mathcal{L}} \chi^2(d-r). \quad (2.13)$$

3. Numerical Studies

3.1. Computation of the estimates

Solving the joint estimating equations (2.1) and (2.6) poses some interesting challenges, since the functions $\hat{g}(\boldsymbol{\beta}^\top \mathbf{X})$ and $\hat{g}'(\boldsymbol{\beta}^\top \mathbf{X})$ depend on $\boldsymbol{\beta}$ implicitly. Treating $\boldsymbol{\beta}^\top \mathbf{X}$ as a new predictor (with given $\boldsymbol{\beta}$), equation (2.1) gives us \hat{g}, \hat{g}' as in Fan, Heckman and Wand (1995). We therefore focus on (2.6), as estimating equations. It cannot be solved explicitly, and hence one needs to find solutions using numerical methods. The Newton-Raphson algorithm is one of the popular and successful methods for finding

roots. However, the computational speed of this algorithm crucially depends on the initial value. We propose therefore a fixed point iterative algorithm that is adapted to higher dimension and insensitive to the initial value. It is worth noting that this algorithm can be implemented in the case that $d > n$, because the resultant procedure only involves one-dimensional nonparametric smoothers, thereby avoiding the data sparsity problem caused by high dimensionality.

Rewrite the estimating functions as $\hat{\mathbf{G}}(\boldsymbol{\beta}) = \mathbf{J}^\top \hat{\mathbf{F}}(\boldsymbol{\beta})$ with $\hat{\mathbf{F}}(\boldsymbol{\beta}) = (\hat{F}_1(\boldsymbol{\beta}), \dots, \hat{F}_d(\boldsymbol{\beta}))^\top$ and

$$\hat{F}_s(\boldsymbol{\beta}) = \sum_{i=1}^n \{X_{si} - \hat{h}_s(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \mu' \{ \hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i) \} \hat{g}'(\boldsymbol{\beta}^\top \mathbf{X}_i) V^{-1} \{ \hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i) \} [Y_i - \mu \{ \hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i) \}].$$

Setting $\hat{\mathbf{G}}(\boldsymbol{\beta}) = 0$ and after some simple calculations, we can get that

$$\begin{cases} \beta_1 = |\hat{F}_1(\boldsymbol{\beta})| / \|\hat{\mathbf{F}}(\boldsymbol{\beta})\|, & s = 1 \\ \beta_s^2 = \hat{F}_s^2(\boldsymbol{\beta}) / \|\hat{\mathbf{F}}(\boldsymbol{\beta})\|^2, & s \geq 2 \end{cases} \quad (3.1)$$

Based on this, the fixed point iterative algorithm is summarized as:

Step 0. Choose initial values for $\boldsymbol{\beta}$, denoted by $\boldsymbol{\beta}_{old}$.

Step 1. Solving the estimating equation (2.1) with respect to $\boldsymbol{\alpha}$, yields $\hat{g}(\boldsymbol{\beta}_{old}^\top \mathbf{X}_i)$ and $\hat{g}'(\boldsymbol{\beta}_{old}^\top \mathbf{X}_i)$, $1 \leq i \leq n$.

Step 2. Update $\boldsymbol{\beta}_{old}$ with $\boldsymbol{\beta}_{old} = \boldsymbol{\beta}_{new} / \|\boldsymbol{\beta}_{new}\|$ by solving the equation (2.6) in the fixed point iteration

$$\boldsymbol{\beta}_{new} = \frac{M}{\hat{F}_1(\boldsymbol{\beta}_{old}) / \|\hat{\mathbf{F}}(\boldsymbol{\beta}_{old})\| + M} \boldsymbol{\beta}_{old} + \frac{|\hat{F}_1(\boldsymbol{\beta}_{old})| / \|\hat{\mathbf{F}}(\boldsymbol{\beta}_{old})\|^2}{\hat{F}_1(\boldsymbol{\beta}_{old}) / \|\hat{\mathbf{F}}(\boldsymbol{\beta}_{old})\| + M} \hat{\mathbf{F}}(\boldsymbol{\beta}_{old}),$$

where M is a constant satisfying $\hat{F}_1(\boldsymbol{\beta}) / \|\hat{\mathbf{F}}(\boldsymbol{\beta})\| + M \neq 0$ for any $\boldsymbol{\beta}$ (M can be chosen to be 2).

Step 3. Repeat Steps 1 and 2 until $\max_{1 \leq s \leq d} |\beta_{new,s} - \beta_{old,s}| \leq tol$ is met with tol being a prescribed tolerance.

The final vector $\boldsymbol{\beta}_{new} / \|\boldsymbol{\beta}_{new}\|$ is the estimator of $\boldsymbol{\beta}^0$. Similar to other direct estimation methods (Horowitz and Härdle, 1996), the preceding calculation is easy to implement. Empirically the initial value for $\boldsymbol{\beta}$, $(1, 1, \dots, 1)^\top / \sqrt{d}$ can be used in the calculations. The Epanechnikov kernel function $K(t) = 3/4(1 - t^2)I(|t| \leq 1)$ is used.

The bandwidth involved in Step 1 can be chosen to be optimal for estimation of $\hat{g}(t)$ and $\hat{g}'(t)$ based on the observations $\{\boldsymbol{\beta}_{old}^\top \mathbf{X}_i, Y_i\}$. So the standard bandwidth selection methods, such as K -fold cross validation, generalized cross validation (GCV) and the rule of thumb, can be adopted. In this step, we recommend K -fold cross validation to determine the optimal bandwidth using the quasi-likelihood as a criterion function. The K -fold cross validation is not too computationally intensive while making K not take too large values (for example, $K = 5$ or 10). Here we recommend that the smoothing parameters be chosen graphically using partial residual plots. The simplest way to do this is to try a number of smoothing parameters that smooth the data and pick the one that seems most reasonable.

3.2. Simulation results

Example 1 (continuous response). We report a simulation study to investigate the finite-sample performance of the proposed estimator and compare it with the rMAVE (refined MAVE, for details see Xia et al. (2002)) estimator. We consider the following model similar to that used in Xia (2006):

$$\begin{aligned} E(Y|\boldsymbol{\beta}^\top \mathbf{X}) &= g(\boldsymbol{\beta}^\top \mathbf{X}), & g(\boldsymbol{\beta}^\top \mathbf{X}) &= (\boldsymbol{\beta}^\top \mathbf{X})^2 \exp(\boldsymbol{\beta}^\top \mathbf{X}); \\ \text{Var}(Y|\boldsymbol{\beta}^\top \mathbf{X}) &= \sigma^2, & \sigma &= 0.1. \end{aligned} \tag{3.2}$$

Let the true parameter $\boldsymbol{\beta} = (2, 1, 0, \dots, 0)^\top / \sqrt{5}$. Two sets of designs for \mathbf{X} are considered: Design (A) and Design (B). In Design (A), $(X_s + 1)/2 \sim \text{Beta}(\tau, 1)$, $1 \leq s \leq d$ and in Design (B) $(X_1 + 1)/2 \sim \text{Beta}(\tau, 1)$ and $P(X_s = \pm 0.5) = 0.5$, $s = 2, 3, 4, \dots, d$. The data generated in Design (A) are not elliptically symmetric. All the components of Design (B) are discrete except for the first component X_1 . Y generates from normal distribution. This simulation data set consists of 400 observations with 250 replications. The results are shown in Table 1. Both MAVE and EFM estimates are close to the true parameter vector for $d = 10$. However, the average estimation errors from MAVE estimates for $d = 50$ are about 10 times larger as those of the EFM estimates. This indicates that the fixed point algorithm is more adaptive to high dimension.

Example 2 (binary response). This simulation design assumes an underlying generalized

single index model for binary responses with

$$\begin{aligned} P(Y = 1|\mathbf{X}) &= \mu\{g(\boldsymbol{\beta}^\top \mathbf{X})\} = \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}/[1 + \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}], \\ g(\boldsymbol{\beta}^\top \mathbf{X}) &= \exp(5\boldsymbol{\beta}^\top \mathbf{X} - 2)/\{1 + \exp(5\boldsymbol{\beta}^\top \mathbf{X} - 3)\} - 1.5. \end{aligned} \quad (3.3)$$

The underlying regression coefficients are assumed to be $\boldsymbol{\beta} = (2, 1, 0, \dots, 0)^\top/\sqrt{5}$. We consider two sets of designs: Design (C) and Design (D). In Design (C), X_1 and X_2 follow the uniform distribution $U(-2, 2)$. In Design (D), X_1 is also assumed to be uniformly distributed in interval $(-2, 2)$ and $(X_2 + 1)/2 \sim \text{Beta}(1, 1)$. The similar designs for generalized partially linear single index models are assumed in Kane, Holt and Allen (2004). Here a sample size of 700 is used for the case $d = 10$ and 3000 is used for $d = 50$. Different sample sizes from Example 1 are used due to varying complexity of the two examples. For this example, 250 replications are simulated and the results are displayed in Table 2. In this set of simulations, the average estimation errors from rMAVE estimates are about three times as large as EFM estimates under both Design (C) and Design (D) for $d = 10$ or $d = 50$. The values in the row marked by $d = 50$ look a little bigger. However, it is reasonable because of the number of summands in the average estimate error for $d = 50$ is five times as large as that for $d = 10$. Again it appears that the EFM procedure achieves more precise estimators.

Table 1: Average estimation errors $\sum_{s=1}^d |\hat{\beta}_s - \beta_s|$ for model (3.2)

d	τ	Design (A)		Design (B)	
		rMAVE	EFM	rMAVE	EFM
10	0.75	0.0559*	0.0809	0.0522*	0.0715
10	1.5	0.0323*	0.0431	0.0417*	0.0523
50	0.75	6.2019	0.9119	6.1265	0.5066
50	1.5	3.1792	0.2313	3.9514	0.2769

* The values are adopted from Xia (2006).

Table 2: Average estimation errors $\sum_{s=1}^d |\hat{\beta}_s - \beta_s|$ for model (3.3)

d	Design (C)		Design (D)	
	rMAVE	EFM	rMAVE	EFM
10	1.2600	0.4785	1.8096	0.7415
50	4.8857	1.2009	5.8280	1.9908

Table 3: Average estimation errors $\sum_{s=1}^d |\hat{\beta}_s - \beta_s|$ for model (3.4)

	$d = 10$	$d = 50$	$d = 100$	$d = 120$
rMAVE	0.6841	6.7224	—	—
EFM	0.0955	0.9363	4.4518	5.9280

— means that the values can not be calculated by rMAVE because of high dimension.

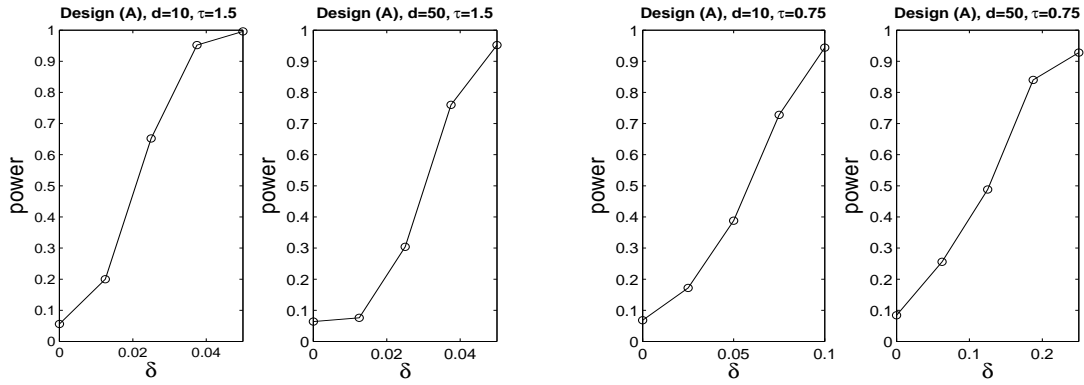


Figure 1. Simulation results for Design (A) in Example 1. The left graphs depict the case $\tau = 1.5$ with τ is the first parameter in $Beta(\tau, 1)$. The right graphs are for $\tau = 0.75$.

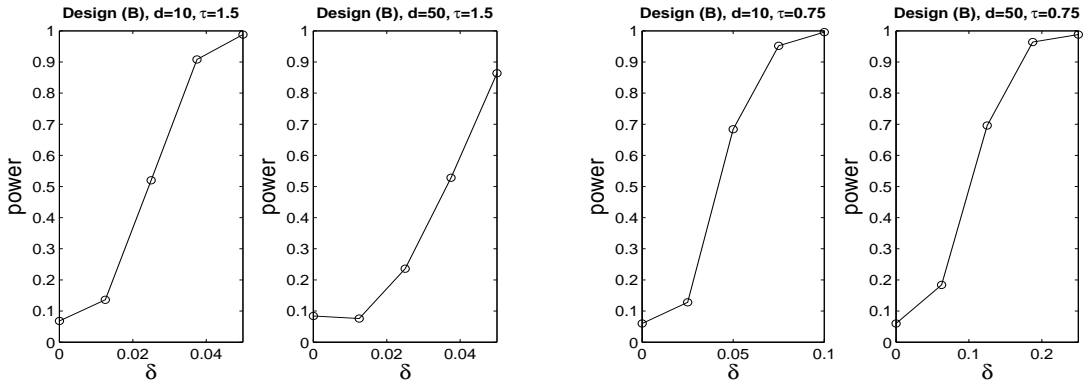


Figure 2. Simulation results for Design (B) in Example 1. The left graphs depict the case $\tau = 1.5$ with τ is the first parameter in $Beta(\tau, 1)$. The right graphs are for $\tau = 0.75$.

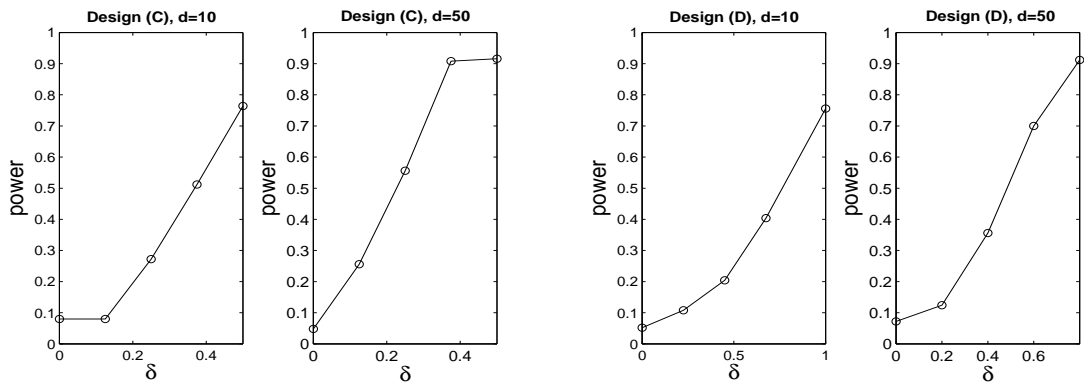


Figure 3. Simulation results for Example 2. The left graphs depict the case of Design (C) with parameter dimension being 10 and 50. The right graphs are for Design (D).

Example 3 (A simple model). To illustrate the adaptivity of our algorithm to high dimension,

we consider the following simple single index model

$$Y = (\boldsymbol{\beta}^\top \mathbf{X})^2 + \varepsilon. \quad (3.4)$$

The true parameter is $\boldsymbol{\beta} = (2, 1, 0, \dots, 0)^\top$, \mathbf{X} is generated from $N_d(0, \mathbf{I})$ and $\varepsilon \sim N(0, 0.2^2)$. The results given in Table 3 are based on a sample size of $n = 100$ and 250 replicates. An important result from this simulation is that the proposed EFM procedure can provide a reasonable estimate even when the dimension of parameter is larger than the number of observations.

Performance of profile quasi-likelihood ratio test. To illustrate how the profile quasi-likelihood ratio performs for linear hypothesis problems, we simulate the same data as above, except that we allow some components of the index to follow the null hypothesis:

$$H_0 : \beta_4 = \beta_5 = \dots = \beta_d = 0.$$

We examine the power of the test under a sequence of the alternative hypotheses indexed by parameter δ as follows:

$$H_1 : \beta_4 = \delta, \quad \beta_s = 0 \quad \text{for } s \geq 5.$$

When $\delta = 0$, the alternative hypothesis becomes the null hypothesis.

We examine the profile quasi-likelihood ratio test under a sequence of alternative models, progressively deviating from the null hypothesis, namely, as δ increases. The power functions are calculated at the significance level: 0.05, using the asymptotic distribution. We calculate test statistics from 250 simulations by employing the fixed point algorithm and find the percentage of test statistics greater than or equal to the associated quantile of the asymptotic distribution. The pictures in Figure 1, 2 and 3 illustrate the power function curves for two models under the given significance levels. The power curves increase rapidly with δ , which shows the profile quasi-likelihood ratio test is powerful. When δ is close to 0, the test sizes are all approximately the significance levels.

3.3. A real data example

Income, to some extent, is considered as an index of a successful life. It is generally believed that demographic information, such as education level, relationship in the household, marital status, the fertility rate and gender, among others, have effects on amounts of income. For example, Murray (1997) illustrated that adults with higher intelligence have higher income.

Kohavi (1996) predicted income using a Bayesian classifier offered by a machine learning algorithm. Madalozzo (2008) examined income differentials between married women and those who remain single or cohabitation by using multivariate linear regression. Here we will use the generalized single index model to explore the relationship between income and some of its possible determinants.

We use the “Adult” database, which was extracted from the census bureau database and is publicly available from website: <http://archive.ics.uci.edu/ml/datasets/Adult>. It was originally used to model income exceeds over USD 50000/year based on census data. After excluding a few missing data, the data set in our study includes 30162 subjects. The selected explanatory variables are:

X_1 = **age** (integer): number of years of age and greater than or equal to 17.

X_2 = **work-class** (categorical): 1 = Federal-gov, 2 = Local-gov, 3 = Never-worked, 4 = Private, 5 = Self-emp-inc (self-employed, incorporated), 6 = Self-emp-not-inc (self-employed, not incorporated), 7 = State-gov, 8 = Without-pay.

X_3 = **fnlwgt** (continuous): The final sampling weights on the CPS files are controlled to independent estimates of the civilian noninstitutional population of the US.

X_4 = **education** (ordinal): 1=Preschool (less than 1st Grade), 2=1st-4th, 3=5th-6th, 4=7th-8th, 5=9th, 6=10th, 7=11th, 8=12th (12th Grade no Diploma), 9=HS-grad (high school Grad-Diploma or Equiv), 10=Some-college (some college but no degree), 11=Assoc-voc (associate degree-occupational/vocationl), 12=Assoc-acdm (associate degree-academic program), 13=Bachelors, 14=Masters, 15=Prof-school (professional school), 16=Doctorate.

X_5 = **education-num** (continuous): Number of years of education.

X_6 = **marital-status** (categorical): 1 = Divorced, 2 = Married-AF-spouse (married, armed forces spouse present), 3 = Married-civ-spouse (married, civilian spouse present), 4 = Married-spouse-absent (married, spouse absent (exc. separated)), 5 = Never-married, 6 = Separated, 7 = Widowed.

X_7 = **occupation** (categorical): 1 = Adm-clerical (administrative support and clerical), 2 = Armed-Forces, 3 = Craft-repair, 4 = Exec-managerial (executive-managerial), 5 = Farming-fishing, 6 = Handlers-cleaners, 7 = Machine-op-inspct (machine operator inspection), 8 = Other-service, 9 = Priv-house-serv (private household services), 10= Prof-specialty (professional specialty), 11= Protective-serv, 12= Sales, 13= Tech-support, 14= Transport-moving.

X_8 = **relationship** (categorical): 1 = Husband, 2 = Not-in-family, 3 = Other-relative, 4 = Own-child, 5 = Unmarried, 6 = Wife.

X_9 = **race** (categorical): 1 = Amer-Indian-Eskimo, 2 = Asian-Pac-Islander, 3 = Black, 4 = Other, 5 = White .

X_{10} = **sex** (categorical): 1=Male, 0=Female.

X_{11} = **capital-gain** (continuous): a profit that results from investments into a capital asset.

X_{12} = **capital-loss** (continuous): a loss that results from investments into a capital asset.

X_{13} = **hours-per-week** (continuous): usual number of hours worked per week.

X_{14} = **native-country** (categorical): 1=United-States, 0=others.

The generalized single index model will be used to model the relationship between income and the relevant 14 predictors $\mathbf{X} = (X_1, \dots, X_{14})^\top$:

$$P(\text{“income”} > 50000|\mathbf{X}) = \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\} / [1 + \exp\{g(\boldsymbol{\beta}^\top \mathbf{X})\}], \quad (3.5)$$

where $Y = I(\text{“income”} > 50000)$ and $\boldsymbol{\beta} = (\beta_1, \dots, \beta_{14})^\top$ and β_s represents the effect of the sth predictor. Formally, we are testing

$$H_0 : \beta_7 = 0 \longleftrightarrow H_1 : \beta_7 \neq 0. \quad (3.6)$$

After a preliminary data check, we find that the explanatory variables X_3 = “fmlwgt”, X_{11} = “capital-gain” and X_{12} = “capital-loss” are very skewed to the left and the later two often take zero value. So before fitting (3.5) we first transformed these three variables. A natural choice is to use $\log(\text{“fmlwgt”})$, $\log(1 + \text{“capital-gain”})$ and $\log(1 + \text{“capital-loss”})$. The fixed point iterative algorithm is employed to compute the estimate for $\boldsymbol{\beta}$. To illustrate further the practical implications of this approach, we compare our results to those obtained by using a logistic regression. The coefficients of the two models are given in Table 4. The relative magnitudes as well as the sign of the estimated coefficients are good proxies of the relevance of each of the explanatory variables. In both models, the variables have the same expected sign. The generalized single index model provides more reasonable results: X_5 = “education-num” has its strongest positive effect on income, those who got a bachelor’s degree or higher seem to have much higher income than those with lower education level.

In contrast, results derived from a logistic regression show that “sex” is the largest positive contributor, an unconvincing conclusion. It appears that men consistently earn more than

women on average — and this is true for lawyers, doctors, etc. Men at most achieved about twice income of their female counterparts in each education level, however, the median weekly earnings of full time workers with doctorate degree is about five times that of those with lowest education (pre-schools). So “education” has a larger impact on income than “sex”.

Some other interesting conclusions could be obtained by looking at the output. “Age” accounts for the experience effect and has a positive effect. Persons who worked without pay in a family business, unpaid childcare and others earn a lower income than persons who worked for wages or for themselves. The “fmlwgt” attribute has a positive relation to income. Males are likely to make much more money (about 1.4 times) than females. The expected sign for marital status is negative, given that the household production theory affirms that division of work is efficient when each member of a family dedicates their time to the more productive job. Men usually receive relatively better compensation for their time in the labor market than in home production. Thus, the expectation is that married women dedicate more time to home tasks and less to the labor market, and this would imply a different probability of working given the marital status choice.

Also “race” influences the income and Asian or Pacific Islanders seem to make more money than other races. And also, one’s income significantly increases as working hours increases. Both “capital-gain” and “capital-loss” have positive effects, so we think that people make more money who can use more money to invest. The presence of young children has a negative influence on the income. The coefficient for “native-country” and “occupation” indicators are not significant. Hence the conclusion based on the generalized single index model is consistent with what we expect. To help with interpretation of the model, plots of $\beta^\top \mathbf{X}$ versus predicted response probability and $g(\beta^\top \mathbf{X})$ are generated, respectively, and can be found on the right column in Figure 4.

We now employ the quasi-likelihood ratio test to the test problem (3.6). The QLR test statistic is 172.4157 with one degree of freedom, resulting in a P-value of $< 10^{-5}$. Hence this result provides strong evidence that occupation has a significant influence on high income.

Table 4: Fitted coefficients for model (3.5)

Variables	$\hat{\beta}$ of GSIM	$\hat{\beta}$ of LR
Constant	—	-9.1710 (.3789)
Age	0.0835 (.0004)	0.0371 (.0014)
Workclass	-0.0664 (.0037)	-0.0867 (.0132)
log(Fnlwgt)	0.0476 (.0055)	0.1104 (.0258)
Education	0.0379 (.0009)	0.0161 (.0053)
Education-Num	0.6517 (.0018)	0.3414 (.0074)
Marital-Status	-0.3163 (.0020)	-0.2163 (.0123)
Occupation	0.0137 (.0009)	0.0002 (.0040)
Relationship	-0.3511 (.0027)	-0.1193 (.0014)
Race	0.0953 (.0036)	0.1035 (.0223)
Sex	0.1792 (.0059)	0.9135 (.0507)
log(1 + Capital-Gain)	0.4569 (.0025)	0.2054 (.0060)
log(1 + Capital-Loss)	0.2290 (.0031)	0.1573 (.0087)
Hours-per-Week	0.0734 (.0004)	0.0297 (.0015)
Native-Country	0.0091 (.0113)	0.0069 (.0656)

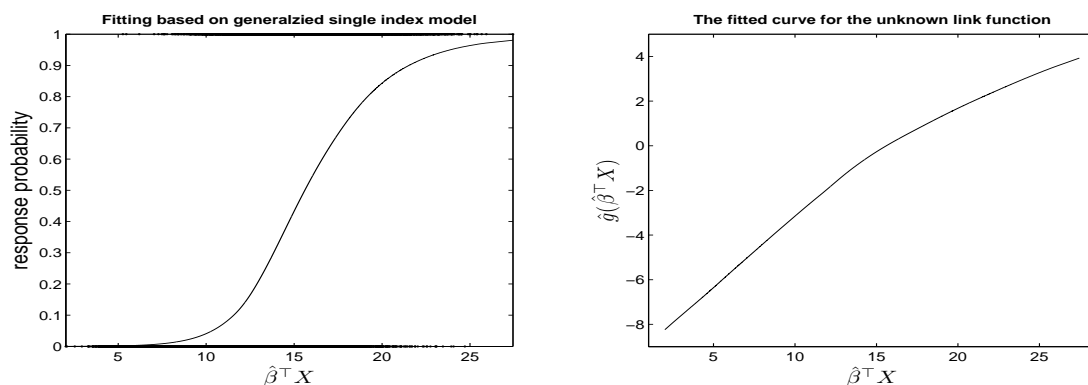


Figure 4. Adult data: The left graph is a plot of predicted response probability based on the generalized single index model. The right graph is the fitted curve for the unknown link function $g(\cdot)$.

References

- CARROLL, R. J., FAN, J., GIJBELS, I. and WAND, M. P. (1997). Generalized partially linear single-index models. *J. Am. Statist. Ass.*, **92** 447–489.
- CARROLL, R. J., RUPPERT, D. and WELSH, A. H. (1998). Local estimating equations. *J. Am. Statist. Ass.*, **93** 214–227.
- FAN, J. and GIJBELS, I. (1996). *Local polynomial modeling and its applications*. London: Chapman and Hall.
- FAN, J., HECKMAN, N. E. and WAND, M. P. (1995). Local polynomial kernel regression for generalized linear models and quasi-likelihood functions. *J. Am. Statist. Ass.*, **90** 141–150.
- FAN, J. and JIANG, J. (2007). Nonparametric inference with generalized likelihood ratio test. *Test*, **16** 409–478.
- HÄRDLE, W., HALL, P. and ICHIMURA, H. (1993). Optimal smoothing in single-index models. *Ann. Statist.*, **21** 157–178.
- HÄRDLE, W. and MAMMEN, E. (1993). Testing parametric versus nonparametric regression. *Ann. Statist.*, **21** 1926–1947.
- HÄRDLE, W., MAMMEN, E. and MÜLLER, M. (1998). Testing parametric versus semiparametric modelling in generalized linear models. *J. Am. Statist. Ass.*, **93** 1461–1474.

- HÄRDLE, W., MAMMEN, E. and PROENCA, I. (2001). A bootstrap test for single index models. *Statistics*, **35** 427–452.
- HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by method of average derivatives. *J. Am. Statist. Ass.*, **84** 986–995.
- HOROWITZ, J. L. and HÄRDLE, W. (1996). Direct semiparametric estimation of a single-index model with discrete covariates. *J. Am. Statist. Ass.*, **91** 1632–1640.
- HRISTACHE, M., JUDITSKI, A. and SPOKOINY, V. (2001). Direct estimation of the index coefficients in a single-index model. *Ann. Statist.*, **29** 595–623.
- KANE, M., HOLT, J. and ALLEN, B. (2004). Results concerning the generalized partially linear single-index model. *J. Statist. Computn Simuln.*, **72** 897–912.
- KOHAVI, R. (1996). Scaling up the accuracy of naive-bayes classifiers: a decision-tree hybrid. In *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*. CA: AAAI Press, Menlo Park, 202–207.
- MADALOZZO, R. C. (2008). An analysis of income differentials by marital status. *Estudos Econômicos*, **38** 267–292.
- MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models (2nd ed.)*. London: Chapman and Hall.
- MURRAY, C. (1997). IQ and economic success. *The Public Interest*.
- POWELL, J. L., STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica*, **57** 1403–1430.
- XIA, Y. (2006). Asymptotic distributions for two estimators of the single-index model. *Economet Theory*, **22** 1112–1137.
- XIA, Y., TONG, H., LI, W. K. and ZHU, L. (2002). An adaptive estimation of dimension reduction space (with discussions). *J. R. Statist. Soc. B.*, **64** 363–410.
- ZHU, L. X. and XUE, L. G. (2006). Empirical likelihood confidence regions in a partially linear single-index model. *J. R. Statist. Soc. B.*, **68** 549–570.

4. Appendix

In this section the proof of Theorem 1 will be given. We first introduce some regularity conditions.

Regularity Conditions

- (a) $\mu(\cdot), V(\cdot), g(\cdot), \mathbf{h}(\cdot) = E(\mathbf{X}|\beta^\top \mathbf{X} = \cdot)$ have two bounded and continuous derivatives.
- (b) Let $q(z, y) = \mu'(z)V^{-1}(z)\{y - \mu(z)\}$. Assume that $\partial q(z, y)/\partial z < 0$ for $z \in \mathbb{R}$ and y in the range of the response variable.
- (c) The largest eigenvalues of $\mathbf{\Omega}_{22}$ is bounded away from infinity.
- (d) The density function $f_{\beta^\top \mathbf{X}}(\beta^\top \mathbf{x})$ of random variable $\beta^\top \mathbf{X}$ is bounded away from 0 on T_β and satisfies the Lipschitz condition of order 1 on T_β , where $T_\beta = \{\beta^\top \mathbf{x} : \mathbf{x} \in T\}$ and T is a compact support set of \mathbf{X} .
- (e) The kernel K is a bounded and symmetric density function with a bounded derivative, and satisfies

$$\int_{-\infty}^{\infty} t^2 K(t) dt \neq 0 \quad \text{and} \quad \int_{-\infty}^{\infty} |t|^j K(t) dt < \infty, j = 1, 2, \dots$$

Condition (a) is some mild smoothness conditions on the involved functions of the model. We impose Condition (b) to guarantee that the solutions of equation (2.1), $\hat{g}(t)$ and $\hat{g}'(t)$, lie in a compact set. Condition (c) implies that the second moment of estimating equation (2.7), $tr(\mathbf{J}^\top \mathbf{\Omega} \mathbf{J})$, is bounded. Then the CLT can be applied to $G(\beta)$. Condition (d) means that \mathbf{X} may have discrete components and the density function of $\beta^\top \mathbf{X}$ is positive, which ensures that the denominators involved in the nonparametric estimators, with high probability, bounded away from 0. Condition (e) is a commonly used smoothness condition, including the Gaussian kernel and the quadratic kernel. All of the conditions can be relaxed at the expense of longer proofs.

Throughout the Appendix, $Z_n = \mathcal{O}_P(a_n)$ denotes that $a_n^{-1}Z_n$ is bounded in probability and the derivation for the order of Z_n is based on the fact that $Z_n = \mathcal{O}_P\{\sqrt{E(Z_n^2)}\}$. Therefore, it allows to apply the Cauchy-Schwartz inequality to the quantity having stochastic order a_n .

Proof of Proposition 1.

- (i). Conditions (a), (b), (d) and (e) are essentially equivalent conditions given by Carroll, Ruppert and Welsh (1998), and as a consequence the derivation of bias and variance for

$\hat{g}(\boldsymbol{\beta}^\top \mathbf{x})$ and $\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x})$, is similar to that of Carroll, Ruppert and Welsh (1998). The asymptotic expansion for $\hat{g}(\boldsymbol{\beta}^\top \mathbf{x})$ and $\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x})$ is helpful in the following proofs. The asymptotic expansion is:

$$\begin{aligned} \hat{g}(\boldsymbol{\beta}^\top \mathbf{x}) - g(\boldsymbol{\beta}^\top \mathbf{x}) &= \left[n^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \right]^{-1} \\ &\quad \times n^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) q \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \\ &\quad + \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-2})^{1/2}, \end{aligned} \quad (\text{A.1})$$

and

$$\begin{aligned} &h\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x}) - hg'(\boldsymbol{\beta}^\top \mathbf{x}) \\ &= \left[n^{-1} \sum_{j=1}^n \{ (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) / h \}^2 K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \right]^{-1} \\ &\quad \times n^{-1} \sum_{j=1}^n \{ (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) / h \} K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) q \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \\ &\quad - \left[n^{-1} \sum_{j=1}^n \{ (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) / h \}^2 K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \right]^{-1} \\ &\quad \times \left[n^{-1} \sum_{j=1}^n \{ (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) / h \} K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \right] \\ &\quad \quad \times \left[n^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \right]^{-1} \\ &\quad \quad \times n^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) [\mu \{ g(\boldsymbol{\beta}^\top \mathbf{X}_j) \} - \mu \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \}] \\ &\quad + \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-2})^{1/2}, \end{aligned} \quad (\text{A.2})$$

where $\alpha_0 = g(\boldsymbol{\beta}^\top \mathbf{x})$, $\alpha_1 = g'(\boldsymbol{\beta}^\top \mathbf{x})$, $q(z, y) = \mu'(z)V^{-1}(z)\{y - \mu(z)\}$, $\rho_l(z) = \{\mu'(z)\}^l V^{-1}(z)$ and $\mathcal{O}_P(1)$ denoting a random quantity is bounded in probability.

(ii). The first equation of (2.1) is

$$\begin{aligned} 0 &= \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mu' \{ \hat{\alpha}_0 + \hat{\alpha}_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} V^{-1} \{ \hat{\alpha}_0 + \hat{\alpha}_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \\ &\quad \times [Y_j - \mu \{ \hat{\alpha}_0 + \hat{\alpha}_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \}]. \end{aligned}$$

Taking derivatives with respect to $\boldsymbol{\beta}^{(1)}$ on both sides, direct observations lead to

$$\frac{\partial \hat{\alpha}_0}{\partial \boldsymbol{\beta}^{(1)}} = \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \{A_1(\boldsymbol{\beta}^\top \mathbf{x}) + A_2(\boldsymbol{\beta}^\top \mathbf{x}) + A_3(\boldsymbol{\beta}^\top \mathbf{x})\},$$

where

$$\begin{aligned}
B(\boldsymbol{\beta}^\top \mathbf{x}) &= - \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) q'_z \{ \hat{\alpha}_0 + \hat{\alpha}_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \}, \\
A_1(\boldsymbol{\beta}^\top \mathbf{x}) &= \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top (\mathbf{X}_j - \mathbf{x}) q'_z \{ \hat{\alpha}_0 + \hat{\alpha}_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \hat{\alpha}_1, \\
A_2(\boldsymbol{\beta}^\top \mathbf{x}) &= \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) q'_z \{ \hat{\alpha}_0 + \hat{\alpha}_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \frac{\partial \hat{\alpha}_1}{\partial \boldsymbol{\beta}^{(1)}}, \\
A_3(\boldsymbol{\beta}^\top \mathbf{x}) &= \sum_{j=1}^n h^{-1} K'_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top (\mathbf{X}_j - \mathbf{x}) q \{ \hat{\alpha}_0 + \hat{\alpha}_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \},
\end{aligned}$$

with $K'_h(\cdot) = h^{-1} K'(\cdot/h)$. Note that $\partial \hat{\alpha}_0 / \partial \boldsymbol{\beta}^{(1)} = \partial \hat{g}(\boldsymbol{\beta}^\top \mathbf{x}) / \partial \boldsymbol{\beta}^{(1)}$, then we have

$$\frac{\partial \hat{g}(\boldsymbol{\beta}^\top \mathbf{x})}{\partial \boldsymbol{\beta}^{(1)}} = \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_1(\boldsymbol{\beta}^\top \mathbf{x}) + \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) + \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_3(\boldsymbol{\beta}^\top \mathbf{x}). \quad (\text{A.3})$$

In the following, we will prove that

$$E \|\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_1(\boldsymbol{\beta}^\top \mathbf{x}) - g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{\mathbf{x} - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x})\}\|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-3}), \quad (\text{A.4})$$

the second term in (A.3) is of order $\mathcal{O}_P(h^4 + n^{-1}h)$, and the third term is of order $\mathcal{O}_P(h^4 + n^{-1}h^{-3})$. The combination of (A.3) and these three results can directly lead to Result (ii) of Proposition 1. The proof is summarized in three steps.

Step 1. The analysis of term $\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_1(\boldsymbol{\beta}^\top \mathbf{x})$.

First we analyze $\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_1(\boldsymbol{\beta}^\top \mathbf{x})$, which can be decomposed as follows,

$$\begin{aligned}
\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_1(\boldsymbol{\beta}^\top \mathbf{x}) &= \hat{g}'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{\mathbf{x} - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x})\} \\
&\quad - \hat{g}'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \left[\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j \right. \\
&\quad \left. \times q'_z \{ \hat{\alpha}_0 + \hat{\alpha}_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x}) \right]
\end{aligned} \quad (\text{A.5})$$

By applying the result about $\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x})$ in (i) of Proposition 1 and the classical asymptotic theory of local linear regression estimate for $\hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x})$ (see Fan et al., 1996), it can be shown that

$$E \|\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{\mathbf{x} - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x})\} - g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{\mathbf{x} - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x})\}\|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-3}). \quad (\text{A.6})$$

Next we show that the second moment of the second term in (A.5) is of order $\mathcal{O}_P(h^4 + n^{-1}h^{-1})$.

Define

$$B_1(\boldsymbol{\beta}^\top \mathbf{x}) = \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{ \alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \}. \quad (\text{A.7})$$

By conditions (a), (d) and (e), we obtain several useful results, for $k = 0, 1, 2$,

$$\begin{aligned}
& E \left\| \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})^k \right. \\
& \quad \left. \times q_z'' \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \right\|^2 \\
&= \mathcal{O}_P \{ cI(k=0) + h^4 I(k \neq 0) + (nh)^{-1} h^{2k} \}, \\
& E \left\| \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})^k \right. \\
& \quad \left. \times \rho_1 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} [Y_j - \mu \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \}] \right\|^2 \\
&= \mathcal{O}_P \{ h^4 I(k=0) + h^8 I(k \neq 0) + (nh)^{-1} h^{2k} \},
\end{aligned} \tag{A.8}$$

and that

$$E \left| \{B(\boldsymbol{\beta}^\top \mathbf{x})/n\}^{-1} - \{B_1(\boldsymbol{\beta}^\top \mathbf{x})/n\}^{-1} \right|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-1}). \tag{A.9}$$

It is easy to check that (A.8) also holds for the case where q_z'' is replaced with q_z' , the case ρ_1 is replaced with ρ_1' or ρ_1'' and the case when multiplier \mathbf{X}_j (not \mathbf{X}_j in linear index $\boldsymbol{\beta}^\top \mathbf{X}_j$) is deleted.

Applying a Taylor expansion to $q_z' \{ \hat{\alpha}_0 + \hat{\alpha}_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \}$ at α_0, α_1 , then the second term in (A.5) can be rewritten as

$$R_1 + R_2 + R_3 + R_4 + \text{others}, \tag{A.10}$$

where

$$\begin{aligned}
R_1(\boldsymbol{\beta}^\top \mathbf{x}) &= g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \left[\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j \right. \\
& \quad \left. \times q_z' \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x}) \right], \\
R_2(\boldsymbol{\beta}^\top \mathbf{x}) &= -g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{ \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x}) - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x}) \}, \\
R_3(\boldsymbol{\beta}^\top \mathbf{x}) &= -g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j \\
& \quad \times [\rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} + \rho_2' \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \}] (\hat{\alpha}_0 - \alpha_0), \\
R_4(\boldsymbol{\beta}^\top \mathbf{x}) &= g'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top [\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} - \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}] \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j \\
& \quad \times \rho_2 \{ \alpha_0 + \alpha_1 (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \},
\end{aligned}$$

and “others” = $\mathcal{O}_P(h^4 + n^{-1}h^{-1})$ according to (A.8), (A.9) and Result (i) of Proposition 1. So the second term in (A.5) is dominated by $R_1 + R_2 + R_3 + R_4$, and the conditional second moment of every term is of order $\mathcal{O}_P(h^4 + n^{-1}h^{-1})$. Combining this and the result of (A.6), the proof for this step completes.

This step yields the following result

$$\begin{aligned} \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_1(\boldsymbol{\beta}^\top \mathbf{x}) &= \hat{g}'(\boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top \{\mathbf{x} - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x})\} \\ &\quad - [R_1(\boldsymbol{\beta}^\top \mathbf{x}) + R_2(\boldsymbol{\beta}^\top \mathbf{x}) + R_3(\boldsymbol{\beta}^\top \mathbf{x}) + R_4(\boldsymbol{\beta}^\top \mathbf{x})] \\ &\quad + \mathcal{O}_P(h^4 + n^{-1}h^{-1}), \end{aligned} \quad (\text{A.11})$$

where $R_1(\boldsymbol{\beta}^\top \mathbf{x}), R_2(\boldsymbol{\beta}^\top \mathbf{x}), R_3(\boldsymbol{\beta}^\top \mathbf{x}), R_4(\boldsymbol{\beta}^\top \mathbf{x})$ are introduced in (A.10). $\#$

Step 2. The analysis of term $\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x})$.

This term can be rewritten as,

$$\begin{aligned} &\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) \\ &= \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) + [\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} - \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}] A_2(\boldsymbol{\beta}^\top \mathbf{x}). \end{aligned} \quad (\text{A.12})$$

We first deal with $\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x})$. By a Taylor expansion at α_0, α_1 and using (A.8) and result (i) of Proposition 1, we have that

$$\begin{aligned} &\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) \\ &= R_5(\boldsymbol{\beta}^\top \mathbf{x}) \frac{\partial \hat{\alpha}_1}{\partial \boldsymbol{\beta}^{(1)}} \\ &\quad + \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \frac{\partial \hat{\alpha}_1}{\partial \boldsymbol{\beta}^{(1)}} \\ &\quad \quad \times q_z'' \{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j\} \{(\hat{\alpha}_0 - \alpha_0) + (\hat{\alpha}_1 - \alpha_1)(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\} \\ &\quad + \mathcal{O}_P(h^4 + n^{-1}h^{-1}) \\ &= R_5(\boldsymbol{\beta}^\top \mathbf{x}) \frac{\partial \hat{\alpha}_1}{\partial \boldsymbol{\beta}^{(1)}} + \mathcal{O}_P(h^4 + n^{-1}h^{-1}), \end{aligned} \quad (\text{A.13})$$

where

$$\begin{aligned} R_5(\boldsymbol{\beta}^\top \mathbf{x}) &= -\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2 \{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\} \\ &\quad \times (\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}). \end{aligned}$$

From (A.13) and (A.8), we conclude that

$$E \|\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x})\|^2 = \mathcal{O}_P(h^4 + n^{-1}h).$$

In addition, we have with Cauchy-Schwarz:

$$\begin{aligned} &E \left\| [\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} - \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}] A_2(\boldsymbol{\beta}^\top \mathbf{x}) \right\|^2 \\ &\leq E^{1/2} \left\| [\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} - \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}] \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\} \right\|^4 E^{1/2} \left\| \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) \right\|^4 \\ &= \mathcal{O}_P(h^8 + n^{-1}h^3). \end{aligned}$$

Then it follows that

$$\begin{aligned}
E \left\| \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) \right\|^2 &\leq 2E \left\| \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) \right\|^2 \\
&\quad + 2E \left\| [\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} - \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}] A_2(\boldsymbol{\beta}^\top \mathbf{x}) \right\|^2 \\
&= \mathcal{O}_P(h^4 + n^{-1}h) + \mathcal{O}_P(h^8 + n^{-1}h^3) = \mathcal{O}_P(h^4 + n^{-1}h).
\end{aligned} \tag{A.14}$$

This step gives that

$$\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x}) = R_5(\boldsymbol{\beta}^\top \mathbf{x}) \frac{\partial \hat{\alpha}_1}{\partial \boldsymbol{\beta}^{(1)}} + \mathcal{O}_P(h^4 + n^{-1}h^{-1}), \tag{A.15}$$

with $R_5(\boldsymbol{\beta}^\top \mathbf{x})$ given in (A.13). ‡

Step 3. The analysis of term $\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_3(\boldsymbol{\beta}^\top \mathbf{x})$.

We now proceed to show that this term is of order $\mathcal{O}_P(h^4 + n^{-1}h^{-3})$ and write

$$\begin{aligned}
&\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_3(\boldsymbol{\beta}^\top \mathbf{x}) \\
&= \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_3(\boldsymbol{\beta}^\top \mathbf{x}) + [\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} - \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}] A_3(\boldsymbol{\beta}^\top \mathbf{x}).
\end{aligned} \tag{A.16}$$

Noting that $\int t^k K'(t) dt = 0$ when k is an even number and following similar derivations as used for $\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{x})$, we have that

$$\begin{aligned}
&\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} A_3(\boldsymbol{\beta}^\top \mathbf{x}) \\
&= R_6(\boldsymbol{\beta}^\top \mathbf{x}) \\
&\quad + \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n h^{-1} K'_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top(\mathbf{X}_j - \mathbf{x}) q'_z \{ \alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \\
&\hspace{25em} \times (\hat{\alpha}_0 - \alpha_0) \\
&\quad + \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n h^{-1} K'_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top(\mathbf{X}_j - \mathbf{x}) q'_z \{ \alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \\
&\hspace{25em} \times \{ (\hat{\alpha}_1 - \alpha_1)(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} \\
&\quad + \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-4})^{1/2} \\
&= R_6(\boldsymbol{\beta}^\top \mathbf{x}) + R_7(\boldsymbol{\beta}^\top \mathbf{x}) + \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-4})^{1/2},
\end{aligned} \tag{A.17}$$

where

$$\begin{aligned}
R_6(\boldsymbol{\beta}^\top \mathbf{x}) &= \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n h^{-1} K'_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top(\mathbf{X}_j - \mathbf{x}) \\
&\hspace{15em} \times q \{ \alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j \} \\
R_7(\boldsymbol{\beta}^\top \mathbf{x}) &= -\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n h^{-1} K'_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{J}^\top(\mathbf{X}_j - \mathbf{x}) \\
&\hspace{15em} \times \rho_2 \{ \alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \} (\hat{\alpha}_0 - \alpha_0).
\end{aligned}$$

The last equality in (A.17) is achieved by plugging in the asymptotic expansion for $\hat{\alpha}_0 - \alpha_0$ provided in (A.1) and by merging the terms of order $\mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-4})^{1/2}$ or of smaller order than this one. Similar to the treat on (A.8), we obtain that $E\|R_6(\boldsymbol{\beta}^\top \mathbf{x})\|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-3})$ and $E\|R_7(\boldsymbol{\beta}^\top \mathbf{x})\|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-1})$, therefore, $E\|\{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}A_3(\boldsymbol{\beta}^\top \mathbf{x})\|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-3})$. An application of Cauchy-Schwartz inequality indicates that the second term in (A.16) is ignorable compared to the first term. Therefore

$$E\|\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}A_3(\boldsymbol{\beta}^\top \mathbf{x})\|^2 = \mathcal{O}_P(h^4 + n^{-1}h^{-3}). \quad (\text{A.18})$$

This step indicates that

$$\{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}A_3(\boldsymbol{\beta}^\top \mathbf{x}) = R_6(\boldsymbol{\beta}^\top \mathbf{x}) + R_7(\boldsymbol{\beta}^\top \mathbf{x}) + \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-4})^{1/2}, \quad (\text{A.19})$$

with $R_6(\boldsymbol{\beta}^\top \mathbf{x})$ and $R_7(\boldsymbol{\beta}^\top \mathbf{x})$ are defined in (A.17). $\#$

Combining the above three steps, we obtain Result (ii) of Proposition 1. \square

Remark 3. *By mimicking the forgoing analysis for $\partial\hat{g}(\boldsymbol{\beta}^\top \mathbf{x})/\partial\boldsymbol{\beta}^{(1)}$, we have that*

$$\begin{aligned} & E\left\|\partial\hat{g}'(\boldsymbol{\beta}^\top \mathbf{x})/\partial\boldsymbol{\beta}^{(1)} - g'(\boldsymbol{\beta}^\top \mathbf{x})\mathbf{J}^\top\{\mathbf{x} - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x}) - \mathbf{h}'(\boldsymbol{\beta}^\top \mathbf{x})\} - S_1(\boldsymbol{\beta}^\top \mathbf{x}) - S_2(\boldsymbol{\beta}^\top \mathbf{x})\right\|^2 \\ & = \mathcal{O}_P(h^4 + n^{-1}h^{-3}), \end{aligned}$$

where

$$\begin{aligned} S_1(\boldsymbol{\beta}^\top \mathbf{x}) & = \left[n^{-1} \sum_{j=1}^n \{(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})/h\}^2 K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2\{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\} \right]^{-1} \\ & \quad \times n^{-1} \sum_{j=1}^n h^{-2} \mathbf{J}^\top(\mathbf{X}_j - \mathbf{x}) K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) q\{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}), Y_j\}, \\ S_2(\boldsymbol{\beta}^\top \mathbf{x}) & = \left[n^{-1} \sum_{j=1}^n \{(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})/h\}^2 K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \rho_2[\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})] \right]^{-1} \\ & \quad \times n^{-1} \sum_{j=1}^n h^{-2} \{(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})/h\} \mathbf{J}^\top(\mathbf{X}_j - \mathbf{x}) K_h'(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \\ & \quad \times \rho_1'\{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\} [Y_j - \mu\{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\}]. \end{aligned}$$

Using this, we can conclude the result of step 2 as

$$\begin{aligned} \{B(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1}A_2(\boldsymbol{\beta}^\top \mathbf{x}) & = R_5(\boldsymbol{\beta}^\top \mathbf{x})g''(\boldsymbol{\beta}^\top \mathbf{x})\mathbf{J}^\top\{\mathbf{x} - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x})\} \\ & \quad + R_5(\boldsymbol{\beta}^\top \mathbf{x})\{S_1(\boldsymbol{\beta}^\top \mathbf{x}) + S_2(\boldsymbol{\beta}^\top \mathbf{x})\} + \mathcal{O}_P(h^8 + n^{-1}h)^{1/2}. \end{aligned} \quad (\text{A.20})$$

In a summary, it follows from the three steps and (A.20) that

$$\begin{aligned} \partial\hat{g}(\boldsymbol{\beta}^\top \mathbf{x})/\partial\boldsymbol{\beta}^{(1)} & = \hat{g}'(\boldsymbol{\beta}^\top \mathbf{x})\mathbf{J}^\top\{\mathbf{x} - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{x})\} \\ & \quad - \{R_1(\boldsymbol{\beta}^\top \mathbf{x}) + R_2(\boldsymbol{\beta}^\top \mathbf{x}) + R_3(\boldsymbol{\beta}^\top \mathbf{x}) + R_4(\boldsymbol{\beta}^\top \mathbf{x})\} \\ & \quad + R_5(\boldsymbol{\beta}^\top \mathbf{x})g''(\boldsymbol{\beta}^\top \mathbf{x})\mathbf{J}^\top\{\mathbf{x} - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{x})\} \\ & \quad + R_5(\boldsymbol{\beta}^\top \mathbf{x})\{S_1(\boldsymbol{\beta}^\top \mathbf{x}) + S_2(\boldsymbol{\beta}^\top \mathbf{x})\} \\ & \quad + R_6(\boldsymbol{\beta}^\top \mathbf{x}) + R_7(\boldsymbol{\beta}^\top \mathbf{x}) + \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-4})^{1/2}, \end{aligned} \quad (\text{A.21})$$

where $R_k, k \leq 4$ are given in step 1, R_5 in step 2 and $R_k, k \geq 6$ in step 3.

Proof of (2.6) and (2.7).

Proof of (2.6). We only need to prove that

$$\hat{G}(\boldsymbol{\beta}) - \sum_{i=1}^n \frac{\partial \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}}{\partial \boldsymbol{\beta}^{(1)}} V^{-1}\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] = \mathcal{O}_P(\sqrt{n}), \quad (\text{A.22})$$

and

$$\frac{\partial \hat{G}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}} + \sum_{i=1}^n \frac{\partial \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}}{\partial \boldsymbol{\beta}^{(1)}} V^{-1}\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \left[\frac{\partial \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}}{\partial \boldsymbol{\beta}^{(1)}} \right]^\top = \mathcal{O}_P(n). \quad (\text{A.23})$$

Using the notations introduced in (A.3), the term in the left hand side of (A.22) can be decomposed into the following three terms:

$$\begin{aligned} & \hat{G}(\boldsymbol{\beta}) - \sum_{i=1}^n [\partial \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} / \partial \boldsymbol{\beta}^{(1)}] V^{-1}\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \\ &= C_1 + C_2 + C_3, \end{aligned}$$

where

$$\begin{aligned} C_1 &= \sum_{i=1}^n \mu'\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1}\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \hat{g}'(\boldsymbol{\beta}^\top \mathbf{X}_i) \\ &\quad \times \mathbf{J}^\top \left[\{B(\boldsymbol{\beta}^\top \mathbf{X}_i)\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{X}_i) \mathbf{X}_j \right. \\ &\quad \left. \times q_2'\{\hat{\alpha}_0 + \hat{\alpha}_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{X}_i), Y_j\} - \hat{\mathbf{h}}(\boldsymbol{\beta}^\top \mathbf{X}_i) \right] \\ C_2 &= - \sum_{i=1}^n \mu'\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1}\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \{B(\boldsymbol{\beta}^\top \mathbf{X}_i)\}^{-1} A_2(\boldsymbol{\beta}^\top \mathbf{X}_i) \\ C_3 &= - \sum_{i=1}^n \mu'\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1}\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \{B(\boldsymbol{\beta}^\top \mathbf{X}_i)\}^{-1} A_3(\boldsymbol{\beta}^\top \mathbf{X}_i). \end{aligned}$$

Next we will prove that $C_k = \mathcal{O}_P(\sqrt{n}), k = 1, 2, 3$. Under the assumptions on h in Theorem 1, by Result (ii) of Proposition 1, results given in (A.11), (A.15) and (A.19) and that for any function $\Delta(\cdot)$ defined on $\mathcal{X} = \{\mathbf{X}_i, 1 \leq i \leq n\}$

$$\begin{aligned} & \sum_{i=1}^n \mu'\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1}\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \Delta(\mathcal{X}) = \mathcal{O}_P(\sqrt{n}), \\ & \sum_{i=1}^n \mu'\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1}\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \\ & \quad \times \mathcal{O}_P(h^8 + n^{-1}h + n^{-2}h^{-4})^{1/2} = \mathcal{O}_P(\sqrt{n}), \end{aligned}$$

we observe that it is sufficient to show that $C_k^* = \mathcal{O}_P(\sqrt{n})$, $k = 1, 2, 3$, where

$$\begin{aligned}
C_1^* &= \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \\
&\quad \times \{R_1^*(\boldsymbol{\beta}^\top \mathbf{X}_i) + R_3(\boldsymbol{\beta}^\top \mathbf{X}_i) + R_4(\boldsymbol{\beta}^\top \mathbf{X}_i)\}, \\
C_2^* &= - \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \\
&\quad \times R_5(\boldsymbol{\beta}^\top \mathbf{X}_i) \{S_1(\boldsymbol{\beta}^\top \mathbf{X}_i) + S_2(\boldsymbol{\beta}^\top \mathbf{X}_i)\}, \\
C_3^* &= - \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] \\
&\quad \times \{R_6(\boldsymbol{\beta}^\top \mathbf{X}_i) + R_7(\boldsymbol{\beta}^\top \mathbf{X}_i)\}.
\end{aligned} \tag{A.24}$$

with

$$\begin{aligned}
R_1^*(\boldsymbol{\beta}^\top \mathbf{x}) &= \{B_1(\boldsymbol{\beta}^\top \mathbf{x})\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j \rho_1' \{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\} \\
&\quad \times [Y_j - \mu \{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\}].
\end{aligned}$$

We first show that $C_1^* = \mathcal{O}_P(\sqrt{n})$, where

$$\begin{aligned}
C_1^* &= \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] R_1^*(\boldsymbol{\beta}^\top \mathbf{X}_i) \\
&\quad + \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] R_3(\boldsymbol{\beta}^\top \mathbf{X}_i) \\
&\quad + \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] R_4(\boldsymbol{\beta}^\top \mathbf{X}_i)
\end{aligned}$$

The second moment of the first term in the above expression is

$$E \left\| \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] R_1^*(\boldsymbol{\beta}^\top \mathbf{X}_i) \right\|^2 = 1/h, \tag{A.25}$$

and the first term is of order $\mathcal{O}_P(\sqrt{n})$ if $nh \rightarrow \infty$. As a consequence of Cauchy-Schwartz inequality, the second term satisfies

$$\begin{aligned}
&\left| \sum_{i=1}^n \mu' \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}] R_3(\boldsymbol{\beta}^\top \mathbf{X}_i) \right|^2 \\
&\leq \sum_{i=1}^n \mu'^2 \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} V^{-2} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}]^2 g'^2(\boldsymbol{\beta}^\top \mathbf{X}_i) \\
&\quad \times \left\| \{B_1(\boldsymbol{\beta}^\top \mathbf{X}_i)\}^{-1} \sum_{j=1}^n K_h(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x}) \mathbf{X}_j \right. \\
&\quad \left. \times [\rho_2 \{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\} + \rho_2' \{\alpha_0 + \alpha_1(\boldsymbol{\beta}^\top \mathbf{X}_j - \boldsymbol{\beta}^\top \mathbf{x})\}] \right\|^2 \\
&\quad \times \sum_{i=1}^n (\hat{\alpha}_0(\boldsymbol{\beta}^\top \mathbf{X}_i) - \alpha_0(\boldsymbol{\beta}^\top \mathbf{X}_i))^2 \\
&= n \mathcal{O}_P(h^2 + n^{-1}h^{-1}) \mathcal{O}_P(h^4 + n^{-1}h^{-1}).
\end{aligned} \tag{A.26}$$

So the second term is of order $\mathcal{O}_P(\sqrt{n})$ when $nh^6 \rightarrow 0$ and $nh^2 \rightarrow \infty$. The structure of R_4, R_7 is similar to that of R_3 and R_5S_1, R_5S_2, R_6 similar to R_1^* respectively, hence we can

use the similar arguments of (A.26) and (A.25) to analyze the remaining terms. Under the bandwidth conditions $nh^6 \rightarrow 0$ and $nh^4 \rightarrow \infty$, we find that $C_k^* = o_P(\sqrt{n})$, $k = 1, 2, 3$. The proof for (A.22) completes. $\#$

(A.23) is a direct consequence of Proposition 1. As a byproduct, we can establish that

$$\begin{aligned} \partial \hat{G}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^{(1)} + \sum_{i=1}^n \{g'(\boldsymbol{\beta}^\top \mathbf{X}_i)\}^2 [\mu'\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}]^2 V^{-1} \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \\ \times \mathbf{J}^\top \{\mathbf{X}_i - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \{\mathbf{X}_i - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}^\top \mathbf{J} = o_P(n). \end{aligned} \quad (\text{A.27})$$

The proof for (2.7). Equation (A.27) implies that $\partial \hat{G}(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^{(1)} - \partial G(\boldsymbol{\beta}) / \partial \boldsymbol{\beta}^{(1)} = o_P(\sqrt{n})$, so we only need concentrate on showing that $\hat{G}(\boldsymbol{\beta}) - G(\boldsymbol{\beta}) = o_P(\sqrt{n})$. After some tedious calculations, we find that it is equivalent to show that the following two dominated terms D_1, D_2 satisfy that $D_k = o_P(\sqrt{n})$, $k = 1, 2$

$$\begin{aligned} D_1 &= \sum_{i=1}^n \mathbf{J}^\top \{\hat{g}'(\boldsymbol{\beta}^\top \mathbf{X}_i) - g'(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \{\mathbf{X}_i - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \rho_1 \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [Y_i - \mu\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\}], \\ D_2 &= \sum_{i=1}^n \mathbf{J}^\top g'(\boldsymbol{\beta}^\top \mathbf{X}_i) \{\mathbf{X}_i - \mathbf{h}(\boldsymbol{\beta}^\top \mathbf{X}_i)\} \rho_1 \{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} [\mu\{g(\boldsymbol{\beta}^\top \mathbf{X}_i)\} - \mu\{\hat{g}(\boldsymbol{\beta}^\top \mathbf{X}_i)\}]. \end{aligned}$$

Plugging (A.2) and (A.1) into above expression, it is straightforward to obtain the desired result. $\#$

We have completed the proof for (2.6) and (2.7). \square

Proof of Theorem 1. Recall the notation $\mathbf{J}, \boldsymbol{\Omega}$ and $\mathbf{G}(\boldsymbol{\beta})$ introduced in Section 2. By (2.7), we have shown that

$$\sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \boldsymbol{\beta}^{(1)0}) = \frac{1}{\sqrt{n}} \{\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J}\}^+ \mathbf{G}(\boldsymbol{\beta}) + o_P(1). \quad (\text{A.28})$$

Theorem 1 follows directly from the above asymptotic expansion and the fact that $E\{\mathbf{G}(\boldsymbol{\beta}) \mathbf{G}^\top(\boldsymbol{\beta})\} = n \mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J}$. \square

Proof of Corollary 1.

The asymptotic covariance of $\hat{\boldsymbol{\beta}}$ can be obtained by adjusting the asymptotic covariance of $\hat{\boldsymbol{\beta}}^{(1)}$ via the multivariate delta method, and is of form $\mathbf{J}(\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^+ \mathbf{J}^\top$. Next we will compare this asymptotic covariance with that (denoted by $\boldsymbol{\Omega}^+$) given in Carroll et al. (1997). Write $\boldsymbol{\Omega}$ as

$$\boldsymbol{\Omega} = \begin{pmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{21} & \boldsymbol{\Omega}_{22} \end{pmatrix}$$

where Ω_{22} is a $(d-1) \times (d-1)$ matrix. We will next investigate two cases respectively: $\det(\Omega_{22}) \neq 0$ and $\det(\Omega_{22}) = 0$. Let $\alpha = -\beta^{(1)}/\sqrt{1 - \|\beta^{(1)}\|^2} = -\beta^{(1)}/\beta_1$.

Consider the case that $\det(\Omega_{22}) \neq 0$. Because $\text{rank}(\Omega) = d-1$, $\det(\Omega_{11}\Omega_{22} - \Omega_{21}\Omega_{12}) = 0$. Note that Ω_{22} is nondegenerate, it can be easily shown that $\Omega_{11} = \Omega_{12}\Omega_{22}^{-1}\Omega_{21}$. Combining this with the following fact,

$$\begin{aligned} \mathbf{J}^\top \Omega \mathbf{J} &= \begin{pmatrix} \alpha & \mathbf{I}_{d-1} \end{pmatrix} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} \alpha^\top \\ \mathbf{I}_{d-1} \end{pmatrix} \\ &= \Omega_{22} + (\Omega_{21}/\sqrt{\Omega_{11}} + \sqrt{\Omega_{11}}\alpha)(\Omega_{12}/\sqrt{\Omega_{11}} + \sqrt{\Omega_{11}}\alpha^\top) - \Omega_{21}\Omega_{12}/\Omega_{11}, \end{aligned}$$

we can get that $\mathbf{J}^\top \Omega \mathbf{J}$ is nondegenerate. In this situation, its inverse $(\mathbf{J}^\top \Omega \mathbf{J})^+$ is just the ordinary inverse $(\mathbf{J}^\top \Omega \mathbf{J})^{-1}$. Then $\mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^+ \mathbf{J}^\top = \left\{ \mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^{-1/2} \right\} \left\{ (\mathbf{J}^\top \Omega \mathbf{J})^{-1/2} \mathbf{J}^\top \right\}$, a full-rank decomposition. Then

$$\begin{aligned} \left\{ \mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^+ \mathbf{J}^\top \right\}^+ &= \left\{ \mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^{-1/2} \right\} \\ &\quad \times \left\{ (\mathbf{J}^\top \Omega \mathbf{J})^{-1/2} \mathbf{J}^\top \mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^{-1} \mathbf{J}^\top \mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^{-1/2} \right\}^{-1} \\ &\quad \times \left\{ (\mathbf{J}^\top \Omega \mathbf{J})^{-1/2} \mathbf{J}^\top \right\} \\ &= \mathbf{J}(\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top \Omega \mathbf{J}(\mathbf{J}^\top \mathbf{J})^{-1} \mathbf{J}^\top \\ &= \Omega. \end{aligned}$$

This means that $\mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^+ \mathbf{J}^\top = \Omega^+$.

When $\det(\Omega_{22}) = 0$, we can obtain that

$$\Omega^+ = \begin{pmatrix} 1/\Omega_{11} + \Omega_{12}\Omega_{22.1}^+ \Omega_{21}/\Omega_{11}^2 & -\Omega_{12}\Omega_{22.1}^+/\Omega_{11} \\ -\Omega_{22.1}^+ \Omega_{21}/\Omega_{11} & \Omega_{22.1}^+ \end{pmatrix},$$

with $\Omega_{22.1} = \Omega_{22} - \Omega_{21}\Omega_{12}/\Omega_{11}$. Write $\mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^+ \mathbf{J}^\top$ as

$$\begin{pmatrix} \alpha^\top (\mathbf{J}^\top \Omega \mathbf{J})^+ \alpha & \alpha^\top (\mathbf{J}^\top \Omega \mathbf{J})^+ \\ (\mathbf{J}^\top \Omega \mathbf{J})^+ \alpha & (\mathbf{J}^\top \Omega \mathbf{J})^+ \end{pmatrix}.$$

Note that $\mathbf{J}^\top \Omega \mathbf{J} = \Omega_{22.1} + (\Omega_{21}/\sqrt{\Omega_{11}} + \sqrt{\Omega_{11}}\alpha)(\Omega_{12}/\sqrt{\Omega_{11}} + \sqrt{\Omega_{11}}\alpha^\top)$, so $\mathbf{J}^\top \Omega \mathbf{J} \geq \Omega_{22.1}$. Combining this with $\text{rank}(\Omega_{22}) = d-2$, we have that $(\mathbf{J}^\top \Omega \mathbf{J})^+ \leq \Omega_{22.1}^+$. It is easy to check that $\alpha^\top \Omega_{22.1} = 0$, so $\alpha \perp \text{span}(\Omega_{22.1})$ and $\alpha^\top \Omega_{22.1}^+ \alpha = 0$, and then $\alpha^\top (\mathbf{J}^\top \Omega \mathbf{J})^+ = 0$. In this situation, $\mathbf{J}(\mathbf{J}^\top \Omega \mathbf{J})^+ \mathbf{J}^\top \leq \Omega^+$ and the stick less-than sign holds since $\mathbf{J}^\top \Omega \mathbf{J} \neq \Omega_{22.1}$ and $1/\Omega_{11} > 0$. \square

Proof of Theorem 2.

Under H_0 , we have $\boldsymbol{\beta} = [\mathbf{e} \ \mathbf{B}]^\top (\sqrt{1 - \|\boldsymbol{\omega}^{(1)}\|^2}, \boldsymbol{\omega}^{(1)\top})^\top$ with $\mathbf{e} = (1, 0, \dots, 0)^\top$ is a r -dimensional vector,

$$\mathbf{B} = \begin{pmatrix} \mathbf{0}^\top & 0 \\ \mathbf{I}_{r-1} & \mathbf{0} \end{pmatrix}$$

is a $r \times (d-1)$ matrix and $\boldsymbol{\omega}^{(1)} = (\beta_2, \dots, \beta_r)^\top$ is a $(r-1) \times 1$ vector. Let $\boldsymbol{\omega} = (\sqrt{1 - \|\boldsymbol{\omega}^{(1)}\|^2}, \boldsymbol{\omega}^{(1)\top})^\top$. So under H_0 the estimator is also the local maximizer $\hat{\boldsymbol{\omega}}$ of the problem

$$\hat{Q}([\mathbf{e} \ \mathbf{B}]^\top \hat{\boldsymbol{\omega}}) = \sup_{\|\boldsymbol{\omega}^{(1)}\| < 1} \hat{Q}([\mathbf{e} \ \mathbf{B}]^\top \boldsymbol{\omega}).$$

Expanding $\hat{Q}(\mathbf{B}^\top \hat{\boldsymbol{\omega}})$ at $\hat{\boldsymbol{\beta}}^{(1)}$ by a Taylor's expansion and noting that $\left. \frac{\partial \hat{Q}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)}} \right|_{\boldsymbol{\beta}^{(1)} = \hat{\boldsymbol{\beta}}^{(1)}} = 0$, then $\hat{Q}(\hat{\boldsymbol{\beta}}) - \hat{Q}(\mathbf{B}^\top \hat{\boldsymbol{\omega}}) = T_1 + T_2 + o_P(1)$, where

$$T_1 = -\frac{1}{2} (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}})^\top \left. \frac{\partial^2 \hat{Q}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)} \partial \boldsymbol{\beta}^{(1)\top}} \right|_{\boldsymbol{\beta}^{(1)} = \hat{\boldsymbol{\beta}}^{(1)}} (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}}),$$

$$T_2 = \frac{1}{6} (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}})^\top \frac{\partial \left\{ (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}})^\top \left. \frac{\partial^2 \hat{Q}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)} \partial \boldsymbol{\beta}^{(1)\top}} \right|_{\boldsymbol{\beta}^{(1)} = \hat{\boldsymbol{\beta}}^{(1)}} (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}}) \right\}}{\partial \boldsymbol{\beta}^{(1)}}.$$

Assuming the conditions in Theorem 1 and under the null hypothesis H_0 , it is easy to show that

$$\sqrt{n}(\mathbf{B}^\top \hat{\boldsymbol{\omega}} - \mathbf{B}^\top \boldsymbol{\omega}) = \frac{1}{\sqrt{n}} \mathbf{B}^\top \mathbf{B} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^+ \mathbf{G}(\boldsymbol{\beta}) + o_P(1).$$

Combining this with (A.28), under the null hypothesis H_0 ,

$$\begin{aligned} & \sqrt{n}(\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}}^{(1)}) \\ &= \frac{1}{\sqrt{n}} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \{ \mathbf{I}_{d-1} - (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2} \mathbf{B}^\top \mathbf{B} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \mathbf{G}(\boldsymbol{\beta}) + o_P(1). \end{aligned} \tag{A.29}$$

Since $\frac{1}{\sqrt{n}} \mathbf{G}(\boldsymbol{\beta}) = o_P(1)$, $\left. \frac{\partial^2 \hat{Q}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^{(1)} \partial \boldsymbol{\beta}^{(1)\top}} \right|_{\boldsymbol{\beta}^{(1)}} = -n \mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J} + o_P(n)$ and matrix $\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J}$ has eigenvalues uniformly bounded away from 0 and infinity, we have $\|\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}}^{(1)}\| = o_P(n^{-1/2})$ and then $|T_2| = o_P(1)$. Combining this and (A.29), we have

$$\begin{aligned} \hat{Q}(\hat{\boldsymbol{\beta}}) - \hat{Q}(\mathbf{B}^\top \hat{\boldsymbol{\omega}}) &= \frac{n}{2} (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}}^{(1)})^\top \mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J} (\hat{\boldsymbol{\beta}}^{(1)} - \mathbf{B}^\top \hat{\boldsymbol{\omega}}^{(1)}) \\ &= \frac{n}{2} \mathbf{G}^\top(\boldsymbol{\beta}) (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \mathbf{P} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \mathbf{G}(\boldsymbol{\beta}), \end{aligned}$$

with $\mathbf{P} = \mathbf{I}_{d-1} - (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2} \mathbf{B}^\top \mathbf{B} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+}$. Here \mathbf{P} is idempotent having rank $d - r$, so it can be written as $\mathbf{P} = \mathbf{S}^\top \mathbf{S}$ where \mathbf{S} is a $(d - r) \times (d - 1)$ matrix satisfying $\mathbf{S} \mathbf{S}^\top = \mathbf{I}_{d-r}$. Consequently,

$$\begin{aligned} 2\{\hat{Q}(\hat{\boldsymbol{\beta}}) - \hat{Q}(\mathbf{B}^\top \hat{\boldsymbol{\omega}})\} &= \left(\sqrt{n} \mathbf{S} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \mathbf{G}(\boldsymbol{\beta}) \right)^\top \left(\sqrt{n} \mathbf{S} (\mathbf{J}^\top \boldsymbol{\Omega} \mathbf{J})^{1/2+} \mathbf{G}(\boldsymbol{\beta}) \right) \\ &\xrightarrow{\mathcal{L}} \chi^2(d - r). \end{aligned}$$

□

SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 001 "Implied Market Price of Weather Risk" by Wolfgang Härdle and Brenda López Cabrera, January 2009.
- 002 "On the Systemic Nature of Weather Risk" by Guenther Filler, Martin Odening, Ostap Okhrin and Wei Xu, January 2009.
- 003 "Localized Realized Volatility Modelling" by Ying Chen, Wolfgang Karl Härdle and Uta Pigorsch, January 2009.
- 004 "New recipes for estimating default intensities" by Alexander Baranovski, Carsten von Lieres and André Wilch, January 2009.
- 005 "Panel Cointegration Testing in the Presence of a Time Trend" by Bernd Droge and Deniz Dilan Karaman Örsal, January 2009.
- 006 "Regulatory Risk under Optimal Incentive Regulation" by Roland Strausz, January 2009.
- 007 "Combination of multivariate volatility forecasts" by Alessandra Amendola and Giuseppe Storti, January 2009.
- 008 "Mortality modeling: Lee-Carter and the macroeconomy" by Katja Hanewald, January 2009.
- 009 "Stochastic Population Forecast for Germany and its Consequence for the German Pension System" by Wolfgang Härdle and Alena Mysickova, February 2009.
- 010 "A Microeconomic Explanation of the EPK Paradox" by Wolfgang Härdle, Volker Krätschmer and Rouslan Moro, February 2009.
- 011 "Defending Against Speculative Attacks" by Tijmen Daniëls, Henk Jager and Franc Klaassen, February 2009.
- 012 "On the Existence of the Moments of the Asymptotic Trace Statistic" by Deniz Dilan Karaman Örsal and Bernd Droge, February 2009.
- 013 "CDO Pricing with Copulae" by Barbara Choros, Wolfgang Härdle and Ostap Okhrin, March 2009.
- 014 "Properties of Hierarchical Archimedean Copulas" by Ostap Okhrin, Yarema Okhrin and Wolfgang Schmid, March 2009.
- 015 "Stochastic Mortality, Macroeconomic Risks, and Life Insurer Solvency" by Katja Hanewald, Thomas Post and Helmut Gründl, March 2009.
- 016 "Men, Women, and the Ballot Woman Suffrage in the United States" by Sebastian Braun and Michael Kvasnicka, March 2009.
- 017 "The Importance of Two-Sided Heterogeneity for the Cyclicity of Labour Market Dynamics" by Ronald Bachmann and Peggy David, March 2009.
- 018 "Transparency through Financial Claims with Fingerprints – A Free Market Mechanism for Preventing Mortgage Securitization Induced Financial Crises" by Helmut Gründl and Thomas Post, March 2009.
- 019 "A Joint Analysis of the KOSPI 200 Option and ODAX Option Markets Dynamics" by Ji Cao, Wolfgang Härdle and Julius Mungo, March 2009.
- 020 "Putting Up a Good Fight: The Galí-Monacelli Model versus 'The Six Major Puzzles in International Macroeconomics'", by Stefan Ried, April 2009.
- 021 "Spectral estimation of the fractional order of a Lévy process" by Denis Belomestny, April 2009.
- 022 "Individual Welfare Gains from Deferred Life-Annuities under Stochastic Lee-Carter Mortality" by Thomas Post, April 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



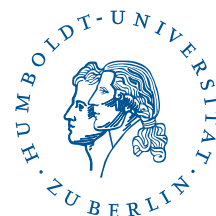
SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 023 "Pricing Bermudan options using regression: optimal rates of convergence for lower estimates" by Denis Belomestny, April 2009.
- 024 "Incorporating the Dynamics of Leverage into Default Prediction" by Gunter Löffler and Alina Maurer, April 2009.
- 025 "Measuring the effects of geographical distance on stock market correlation" by Stefanie Eckel, Gunter Löffler, Alina Maurer and Volker Schmidt, April 2009.
- 026 "Regression methods for stochastic control problems and their convergence analysis" by Denis Belomestny, Anastasia Kolodko and John Schoenmakers, May 2009.
- 027 "Unionisation Structures, Productivity, and Firm Performance" by Sebastian Braun, May 2009.
- 028 "Optimal Smoothing for a Computationally and Statistically Efficient Single Index Estimator" by Yingcun Xia, Wolfgang Härdle and Oliver Linton, May 2009.
- 029 "Controllability and Persistence of Money Market Rates along the Yield Curve: Evidence from the Euro Area" by Ulrike Busch and Dieter Nautz, May 2009.
- 030 "Non-constant Hazard Function and Inflation Dynamics" by Fang Yao, May 2009.
- 031 "De copulis non est disputandum - Copulae: An Overview" by Wolfgang Härdle and Ostap Okhrin, May 2009.
- 032 "Weather-based estimation of wildfire risk" by Joanne Ho and Martin Odening, June 2009.
- 033 "TFP Growth in Old and New Europe" by Michael C. Burda and Battista Severgnini, June 2009.
- 034 "How does entry regulation influence entry into self-employment and occupational mobility?" by Susanne Prantl and Alexandra Spitz-Oener, June 2009.
- 035 "Trade-Off Between Consumption Growth and Inequality: Theory and Evidence for Germany" by Runli Xie, June 2009.
- 036 "Inflation and Growth: New Evidence From a Dynamic Panel Threshold Analysis" by Stephanie Kremer, Alexander Bick and Dieter Nautz, July 2009.
- 037 "The Impact of the European Monetary Union on Inflation Persistence in the Euro Area" by Barbara Meller and Dieter Nautz, July 2009.
- 038 "CDO and HAC" by Barbara Choroś, Wolfgang Härdle and Ostap Okhrin, July 2009.
- 039 "Regulation and Investment in Network Industries: Evidence from European Telecoms" by Michał Grajek and Lars-Hendrik Röller, July 2009.
- 040 "The Political Economy of Regulatory Risk" by Roland Strausz, August 2009.
- 041 "Shape invariant modelling pricing kernels and risk aversion" by Maria Grith, Wolfgang Härdle and Juhyun Park, August 2009.
- 042 "The Cost of Tractability and the Calvo Pricing Assumption" by Fang Yao, September 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".



SFB 649 Discussion Paper Series 2009

For a complete list of Discussion Papers published by the SFB 649, please visit <http://sfb649.wiwi.hu-berlin.de>.

- 043 "Evidence on Unemployment, Market Work and Household Production" by Michael C. Burda and Daniel S. Hamermesh, September 2009.
- 044 "Modelling and Forecasting Liquidity Supply Using Semiparametric Factor Dynamics" by Wolfgang Karl Härdle, Nikolaus Hautsch and Andrija Mihoci, September 2009.
- 045 "Quantifizierbarkeit von Risiken auf Finanzmärkten" by Wolfgang Karl Härdle and Christian Wolfgang Friedrich Kirchner, October 2009.
- 046 "Pricing of Asian temperature risk" by Fred Benth, Wolfgang Karl Härdle and Brenda López Cabrera, October 2009.
- 047 "MM-Stat – MultiMedia-Statistik: Statistische Datenanalyse – webbasiert, interaktiv und multimedial" by Sigbert Klinke, Dina Kuhlee, Christian Theel, Cornelia Wagner and Christian Westermeier, October 2009.
- 048 "Migration of the Highly Skilled: Can Europe catch up with the US?" by Lydia Mechtenberg and Roland Strausz, October 2009.
- 049 "A blocking and regularization approach to high dimensional realized covariance estimation" by Nikolaus Hautsch, Lada M. Kyj and Roel C.A. Oomen, October 2009.
- 050 "Generalized single-index models: The EFM approach" by Xia Cui, Wolfgang Karl Härdle and Lixing Zhu, October 2009.

SFB 649, Spandauer Straße 1, D-10178 Berlin
<http://sfb649.wiwi.hu-berlin.de>

This research was supported by the Deutsche
Forschungsgemeinschaft through the SFB 649 "Economic Risk".

