MONASH University

MONASH BUSINESS SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

http://business.monash.edu/econometrics-and-business-statistics/research/publications

Series Estimation for Single-Index Models under Constraints

Chaohua Dong, Jiti Gao and Bin Peng

April 2018

Working Paper 5/18

Series Estimation for Single–Index Models under Constraints

Chaohua Dong*, Jiti Gao⋆ and Bin Peng†

*Southwestern University of Finance and Economics, ⋆Monash University,
and †University of Bath

16 April, 2018

Abstract

This paper discusses a semiparametric single-index model. The link function is allowed to be unbounded and has unbounded support that fill the gap in the literature. The link function is treated as a point in an infinitely many dimensional function space which enables us to derive the estimates for the index parameter and the link function simultaneously. This approach is different from the profile method commonly used in the literature. The estimator is derived from an optimization with the constraint of an identification condition for the index parameter, which solves an important problem in the literature of single-index models.

In addition, making use of a property of Hermite orthogonal polynomials, an explicit estimator for the index parameter is obtained. Asymptotic properties of the two estimators of the index parameter are established. Their efficiency is discussed in some special cases as well. The finite sample properties of the two estimators are demonstrated through an extensive Monte Carlo study and an empirical example.

Keywords: Asymptotic theory; closed-form estimation; cross-sectional model; Hermite series expansion.

JEL classification: C13, C14, C51

- Chaohua Dong, School of Economics, Southwestern University of Finance and Economics, Chengdu, Sichuan 611130, China. Email: dchaohua@hotmail.com.
- *Correspondence*: Jiti Gao, Department of Econometrics and Business Statistics, Monash University, VIC 3145, Australia. Email: Jiti.Gao@monash.edu.
- Bin Peng, Department of Economics, University of Bath, Bath BA27JP, UK. Email: B.Peng2@bath.ac.uk.

1 Introduction

Single-index models have been studied extensively in the statistics and econometrics literature in the past thirty years or so and cover many classical parametric models by using a general function of the form $g_0(x'\theta_0)$. Among all methods, one important class of estimation methods is based on using a nonparametric kernel method, (e.g., Ichimura (1993), Härdle et al. (1993), Carroll et al. (1997), Xia et al. (1999), Xia et al. (2002), Chapter 2 of Gao (2007), Ma and Song (2015) and Birke et al. (2017), for example). A detailed review on how to employ kernel methods to estimate single-index models can be found in Xia (2006). Alternatively, series–based methods provide good approximations to unknown functions and are of computational convenience (see Chen (2007) for an extensive review). To the best of our knowledge, however, limited studies are available for series–based estimation for single-index models. The existing literature includes Yu and Ruppert (2002), and Ma and Song (2015) for using spline estimation to investigate the cross sectional single-index models, Dong et al. (2015) and Dong et al. (2016) for employing Hermite polynomials to study single-index models in panel data and nonstationary time series models, respectively.

In the literature of single-index models, researchers usually need to assume either the boundedness of the support of the link function or the boundedness of the link function itself. For instance, Assumptons 5.2 and 5.3 of Ichimura (1993) impose compactness on both the parameter space and the support of the regressors, implying that the link function is defined on a compact set. Condition C2 of Xia (2006) and Assumption A of Cai et al. (2015) directly require the boundedness of the link function, although they are defined on the entire real line. More recently, Dong et al. (2016) also require the link function to be smooth and integrable, which basically implies the boundedness of the link function itself.

Moreover, in nonparametric context of using a sieve method, in order to deal with the issue of unboundedness for the support of the regression function under study, Chen and Christensen (2015) truncate the unbounded support by a compact set depending on the sample size. Similar concerns have recently been raised by Hansen (2015), where the author points out that in nonparmetric sieve regression there may only be very limited studies available about the case of unbounded regressors.

Furthermore, we employ an optimization with constraint of the identification condition to derive an estimator of the index parameter. This natural approach is ignored in the literature, and also gives rise to a challenge for asymptotic theory of our estimator as both the estimator and the parameter belong to a proper subspace of the parameter space. By contrast, the existing literature uses a semiparametric least squares estimate, $\overline{\theta}$, for θ_0 and then derives a standardized version of the form $\overline{\theta}/\|\overline{\theta}\|$. Although Ai and Chen (2003) argue that their model and approach may nest the single-index setting as a special case, the constraint of the identification condition is in fact not considered. As can be seen in Theorem 2.2 and Remarks 2 and 3 of this paper, the constraint has a serious impact on both the asymptotic theory and numerical performance, which, to the best of our knowledge, is brought to attention for the first time.

Thus, one of the main contributions of this paper is to deal with the unboundedness of support for the link function in a cross-sectional model of the form:

$$y = g_0(x'\theta_0) + e, \tag{1.1}$$

where the samples $\{y_i, x_i\}_{i=1}^N$ are observable, and for identifiability suppose $\theta_0 = (\theta_{0,1}, \ldots, \theta_{0,d})'$ satisfies $\|\theta_0\| = 1$ with $\theta_{0,1} > 0$ and the link function $g_0(w) \in L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$. Note that Hilbert space $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ covers, to name a few, all polynomials, all power functions and all bounded functions on $\mathbb{R}$ that are often encountered in both applications and econometric theory.[2]

Note also that the function space $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ is of infinite many dimension, where Hermite polynomials form a complete orthonormal basis. From a geometric point of view, the link function $g_0(w)$ has infinitely many coordinates, but the ones in the tail play much less role in the determination of $g_0(w)$. After truncation of the coordinates, $g_0(w)$, along with θ_0, is treated approximately as a finite-dimensional parameter and hence both θ_0 and $g_0(\cdot)$ are estimated simultaneously. This gives the first pair of estimates for $(\theta_0, g_0(\cdot))$.

With the help of Hermite polynomial's property, semiparametric model (1.1) is parameterized completely. This gives the second pair of estimates for $(\theta_0, g_0(\cdot))$ to be studied below. The benefit is two-fold. One is that we are able to obtain a closed-form for the estimator of θ_0; another is that, as in a parametric model case, we need not impose compactness on the parameter space of θ_0. These two improvements being offered by this paper cannot be achieved in the existing literature where θ_0 is estimated based on the assumption of the compactness of the parameter space, see, for example, Dong et al. (2016).

In summary, the main contributions of this paper are given as follows.

- Unlike what has been done in most of the literature, the first estimation method requires neither the boundedness of the support of the regression function nor the boundedness of the regression function itself.
- The estimator derived from an optimization under the constraint $\|\theta_0\| = 1$ may have a super convergence rate $O_P(N^{-1})$ along with the direction of θ_0, while it has root-N rate along with all other directions orthogonal to θ_0.
- The second estimation method offers a closed-form estimator for θ_0 via an ordinary least squares (OLS). As a consequence, the closed-form OLS estimate makes it convenient for both the establishment of an asymptotic theory and the ease of practical implementation.
- We compare the asymptotic efficiency of both methods. Under certain restrictions, it is shown that the methods are asymptotically equivalent.

The organization of this paper is as follows. Section 2 proposes two estimation methods and then establishes the corresponding asymptotic theory. Section 3 evaluates the finite-sample properties for the proposed estimation methods and theory. Section 4 provides an empirical study based on U.S. banking industry. Section 5 concludes the paper with some comments. All the mathematical proofs are given in an appendix.

Throughout this paper, $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ denote the minimum and maximum eigenvalues of a square matrix A, respectively; both M's and $O(1)$'s denote positive constants which may be different at each appearance; $\|\cdot\|$ denotes the Euclidean norm; $\lfloor a \rfloor \le a$ denotes the largest integer not exceeding a. For a given function $g(w)$, its first and second derivatives are denoted by $g^{(1)}(w)$ and $g^{(2)}(w)$, while for notational convenience, sometimes $g^{(0)}(w)$ represents $g(w)$ itself; and when there is no possibility of ambiguity, $\int g(w)dw$ replaces $\int_{\mathbb{R}} g(w)dw$. The same applies to multiple integrals.

[2]Given preassumption of $x'\theta_0 \in \mathbb{R}^+$, Hilbert space $L^2\left(\mathbb{R}^+, \exp(-w)\right)$ for the link function can be adopted. Due to the similarity and inclusion of $\mathbb{R}^+ \subset \mathbb{R}$, only the case of the support of $\mathbb{R}$ is studied in the paper.

2 Estimation Methods

It is known that Hermite polynomials form a complete orthogonal system in the Hilbert space $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ with each element defined by

$$H_m(w) = (-1)^m \cdot \exp(w^2/2) \cdot \frac{d^m}{dw^m} \exp(-w^2/2), \quad m = 0, 1, 2, \ldots \tag{2.1}$$

The orthogonality of this system reads $\int H_m(w)H_n(w)\exp(-w^2/2)dw = m!\sqrt{2\pi}\delta_{mn}$, where δ_{mn} is the Kronecker delta. Further define $h_m(w) = \frac{1}{\sqrt{m!}}H_m(w)$, so that $\{h_m(w)\}$ becomes an orthogonal basis in the Hilbert space $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ satisfying $\int h_m(w)h_n(w)\exp(-w^2/2)dw = \sqrt{2\pi}\delta_{mn}$. Hence, for any $g(w) \in L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$, we have an orthogonal series expansion in terms of $h_m(w)$ as follows:

$$g(w) = \sum_{m=0}^{\infty} c_m h_m(w), \quad \text{with} \quad c_m(g) = \frac{1}{\sqrt{2\pi}} \int g(w)h_m(w)\exp(-w^2/2)dw. \tag{2.2}$$

Define the function norm $\|\cdot\|_{L^2}$ as $\|g\|_{L^2} = \left\{\frac{1}{\sqrt{2\pi}}\int |g(w)|^2 \exp(-w^2/2)dw\right\}^{1/2}$. It follows from Parseval's equality that $\|g\|_{L^2}^2 = \sum_{m=0}^{\infty} c_m^2$. Since the right-hand side is the squared norm of $\{c_m, m = 0, 1, \ldots\}$ in the sequence space ℓ^2, the equality, along with Riesz-Fischer theorem, implies that $\|\cdot\|_{L^2}$ and $\|\cdot\|_{\ell^2}$ are isomorphic, so a function g can be identified by its associated coefficients $\{c_m, m = 0, 1, \ldots\}$. From a geometric point of view, the infinite-dimensional point $g \in L^2(\mathbb{R}, \exp(-w^2/2))$ has coordinates $\{c_m, m = 0, 1, \ldots\}$. Thus, in order to express g, it suffices to determine its coordinate sequence $\{c_m, m = 0, 1, \ldots\}$.

For any truncation parameter $k \geq 1$, split the orthogonal series expansion into two parts,

$$\begin{aligned} g(w) =& g_k(w) + \delta_k(w), \quad \text{where } g_k(w) = \mathcal{H}(w)'C_k \text{ and } \delta_k(w) = \textstyle\sum_{m=k}^{\infty} c_m h_m(w), \\ \mathcal{H}(w) =& (h_0(w), \ldots, h_{k-1}(w))', \quad C_k = (c_0, \ldots, c_{k-1})'. \end{aligned} \tag{2.3}$$

By virtue of (2.3) and Lemma A.3 below, respectively, $g_0(x_i'\theta_0)$ in model (1.1) can have two representations:

$$g_0\left(x_i'\theta_0\right) = g_{0,k}(x_i'\theta_0) + \delta_{0,k}\left(x_i'\theta_0\right), \tag{2.4}$$

$$g_0\left(x_i'\theta_0\right) = \sum_{m=0}^{k-1} \sum_{|\mathsf{p}|=m} a_{m,\mathsf{p}}\left(\theta_0\right) \mathscr{H}_{\mathsf{p}}\left(x_i\right) + \delta_{0,k}\left(x_i'\theta_0\right), \tag{2.5}$$

where equation (2.5) is derived from (2.4) using Lemma A.3,

$$\begin{aligned} & g_{0,k}(x_i'\theta_0) = \mathcal{H}(x_i'\theta_0)'C_{0,k}, \quad \delta_{0,k}(x_i'\theta_0) = \sum_{m=k}^{\infty} c_{0,m}h_m(x_i'\theta_0), \\ & C_{0,k} = (c_{0,0}, \ldots, c_{0,k-1})', \quad c_{0,m} = \frac{1}{\sqrt{2\pi}} \int g_0(w)h_m(w)\exp\left(-w^2/2\right)dw, \\ & a_{m,\mathsf{p}}(\theta_0) = \sqrt{\binom{m}{\mathsf{p}}} c_{0,m}\theta_0^{\mathsf{p}}, \quad \binom{m}{\mathsf{p}} = \frac{m!}{\prod_{j=1}^d p_j!}, \quad \theta_0^{\mathsf{p}} = \prod_{j=1}^d \theta_{0,j}^{p_j}, \\ & \mathscr{H}_{\mathsf{p}}\left(x_i\right) = \prod_{j=1}^d h_{p_j}\left(x_{i,j}\right), \quad x_i = \left(x_{i,1}, \ldots, x_{i,d}\right)', \quad \mathsf{p} = \left(p_1, \ldots, p_d\right)', \end{aligned}$$

$|\mathsf{p}| = p_1 + \ldots + p_d$ and p_j's for $j = 1, \ldots, d$ are non-negative integers.

The expansions (2.4) and (2.5) allow us to use two different methods to estimate θ_0 and g_0 in what follows.

2.1 Extremum Estimation Method

With the expansion (2.4), the nonparametric function g_0 is parameterized by $\{c_m, m = 0, 1, \ldots\}$. Thus, the unknown parameter θ_0 and the nonparametric function g_0 together can be viewed as a point in an infinite-dimensional Euclidean space, the 2-fold Cartesian product space by $\mathbb{R}^d$ and $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$. The space is equipped with norm $\|\cdot\|_2$ given by

$$\|(\theta, g)\|_2 = \left(\|\theta\|^2 + \sum_{m=0}^{\infty} c_m^2(g)\right)^{1/2}, \tag{2.6}$$

where $c_m(g)$'s are the coefficients in the expansion of g and defined in (2.2). Clearly, $\|\cdot\|_2$ satisfies the definition of a norm and is similar to Newey and Powell (2003, p. 1569).

Suppose that $\Theta \subset \mathbb{R}^d$, Θ is compact and $\theta_0 \in \Theta$. Suppose further that G is a subset of $L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ such that $g_0 \in G$ and $\sup_{g\in G} \|g\|_{L^2} < B_1 < \infty$ for some constant B_1.

Model (1.1) implies that $L(\theta, g) := E[y - g(x'\theta)]^2$ attains its minimum at (θ_0, g_0). After taking into account the identification restriction, we introduce the following population version of the objective function with Lagrange multiplier as

$$W_\lambda(\theta, g) = L(\theta, g) + \lambda(\|\theta\|^2 - 1), \tag{2.7}$$

where $(\theta, g) \in \Theta \times G$.

Analogous to (2.7), we denote the corresponding sample version of the objective function as

$$W_{N,\lambda}(\theta, g) = L_N(\theta, g) + \lambda(\|\theta\|^2 - 1), \tag{2.8}$$

where $(\theta, g) \in \Theta \times G_k$, G_k is defined as $G_k = G \cap \mathrm{span}\{h_0(w), h_1(w) \ldots, h_{k-1}(w)\}$, k is the truncation parameter, and $L_N(\theta, g) = \frac{1}{N}\sum_{i=1}^N [y_i - g(x_i'\theta)]^2$. More importantly, an approximate version of (2.8) is given by

$$\mathcal{W}_{N,\lambda}(\theta, C_k) = \mathcal{L}_N(\theta, g) + \lambda(\|\theta\|^2 - 1), \tag{2.9}$$

where $\mathcal{L}_N(\theta, g) = \frac{1}{N}\sum_{i=1}^N [y_i - \mathcal{H}(x_i'\theta)'C_k]^2$, and $\mathcal{H}(w)$ and C_k are defined in (2.3). Mathematically, the estimates of θ_0 and C_k are given by the following minimization:

$$(\widetilde{\theta}, \widetilde{C}_k) = \underset{\Omega_k,\ \lambda}{\operatorname{argmin}}\, \mathcal{W}_{N,\lambda}(\theta, C_k), \tag{2.10}$$

where $\Omega_k = \{(\theta, C_k) : \ \theta \in \Theta, \|C_k\| \leq B_1\} \subseteq \mathbb{R}^{d+k}$. It is noteworthy that (2.10) can easily be implemented by using function "fmincon" in Matlab. Thus, define $\widetilde{g}_k(w) = \widetilde{C}_k'\mathcal{H}(w)$ for any $w \in \mathbb{R}$, and $(\widetilde{\theta}, \widetilde{g}_k(w))$ is the estimate of $(\theta_0, g_0(w))$.

To establish the main results of this paper, we now introduce the following conditions.

Assumption 1:

1. *Suppose that* $\{x_i, e_i\}$ *is independent and identically distributed (i.i.d.) across* $i = 1, \ldots, N$. *Moreover,* $E[e_1|x_1] = 0$ *a.s. and* $E[e_1^2|x_1] = \sigma_e^2$ *a.s., where* σ_e^2 *is a positive constant.*

2. *Suppose* $L(\theta, g) := E[y - g(x'\theta)]^2$ *defined above (2.7) has a unique minimum (i.e.,* σ_e^2*) on* $\Theta \times G$ *at* (θ_0, g_0). *For* $\forall g \in G$, *the first derivative* $g^{(1)}(w)$ *always exists. Moreover, there exists a constant* M *satisfying*

   (a) $\max_{\{(\theta,g)\in\Theta\times G\}} E\left\|x_1 x_1'\{g^{(1)}(x_1'\theta)\}^2\right\| \leq M$;

   (b) $\sup_{\{(\theta,w)\in\Theta\times\mathbb{R}\}} \exp\left(w^2/2\right) f_\theta(w) \leq M$, *where* $f_\theta(w)$ *defines the probability density function (pdf) of* $w = x_1'\theta$ *for each given* $\theta \in \Theta$.

3. (a) *All derivatives* $g_0^{(j)}(w) \in L^2\left(\mathbb{R}, \exp(-w^2/2)\right)$ *for* $j = 0, 1, \ldots, r$ *and* $r \geq 2$;

   (b) *Let* k *be divergent with* N *such that* $k^2/N \to 0$ *as* $N \to \infty$.

Assumption 1.1 can be further relaxed to take into account heteroskedasticity. For example, one can assume $E[e_i^2|x_i] = \sigma^2(x_i)$ for $i = 1, \ldots, N$, and all the proofs will go through with some suitable modification. Moreover, we may also consider a cross-sectional dependence (CSD) case where $e_i = \sigma(x_i)\xi_i$, in which $\{\xi_i, i \geq 1\}$ is a sequence of random errors with $E[\xi_i] = 0$, $E[\xi_i^2] = 1$ and $\gamma_{ij} = E[\xi_i\, \xi_j] \neq 0$ when $i \neq j$. Since establishing the corresponding theory for this cross-sectional dependence setting involves much more techniques than what will be involved in this paper, this paper focuses on the i.i.d. setting, and the CSD case will be left for future research. Assumption 1.2.b rules out heavy-tailed distributions, which is due to the fact that g_0 is potentially unbounded. Many variables such as normal variables and those possessing compact support satisfy this condition. Assumption 1.3 requires the smoothness of the regression function and the divergence restriction of the truncation parameter, but more rigorous requirement of the divergence rate for k will be given in Assumption 2 below so as to the truncation error is negligible. Indeed, by virtue of Lemma A.1, $\|g_0 - g_{0,k}\|_{L^2}^2 = \sum_{m=k}^{\infty} c_{0,m}^2 = O(k^{-r})$ and $\|g_0^{(1)} - g_{0,k}^{(1)}\|_{L^2}^2 = \sum_{m=k}^{\infty} c_{0,m}^2 m = o(1)$ as $k \to \infty$. See Lemma A.1 for the rate of $c_{0,m} \to 0$. Furthermore, in order to simplify the analysis and notation under a general sieve space, the norm provided in Assumption 3 of Newey (1997) can be adopted. Below we explain why Assumptions 1.2.a is reasonable in Remark 1.

Remark 1. *Assumption 1.2.a covers some conditions commonly used in the existing literature as special cases.*

1. *If Assumption 5.3.1 of Ichimura (1993) holds, i.e.,* x *belongs to a compact set, then we can write*

$$\begin{aligned} E\left\|x_1 x_1'\{g^{(1)}(x_1'\theta)\}^2\right\| &\leq O(1) E\left|g^{(1)}(x_1'\theta)\right|^2 = O(1)\int \left\{g^{(1)}(w)\right\}^2 f_\theta(w)dw \\ &\leq O(1)\int \left\{g^{(1)}(w)\right\}^2 \exp(-w^2/2)\cdot\exp(w^2/2) f_\theta(w)dw \\ &\leq O(1)\int \left\{g^{(1)}(w)\right\}^2 \exp(-w^2/2)dw, \end{aligned}$$

   where $f_\theta(w)$ *is the same as that defined in Assumption 1.2.b. Then in this case, Assumption 1.2.a reduces to requiring* $g^{(1)}(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ *for* $\forall g \in G$.

2. If Condition C2 of Xia (2006) holds, i.e., $g^{(1)}(w)$ is bounded on $\mathbb{R}$, then we can write

$$E\left\|x_1x_1'\{g^{(1)}(x_1'\theta)\}^2\right\| \leq O(1)E\|x_1\|^2.$$

Then we need only to bound the second moment of x_1.

Since our link function can potentially be an unbounded function defined on the whole real line, we adopt the current form of Assumption 1.2.a.

With Assumption 1 in hand, we now summarize the consistency of (2.10) below.

Theorem 2.1. *Let Assumption 1 hold. As $N \to \infty$, we have*

$$\left\|(\widetilde{\theta}, \widetilde{g}_k) - (\theta_0, g_0)\right\|_2 \to_P 0,$$

where $\widetilde{g}_k = \widetilde{C}_k'\mathcal{H}(w)$, $\mathcal{H}(w)$ and $\widetilde{C}_k$ are denoted by (2.3) and (2.10), respectively.

We now move on to establish an asymptotic normality. Before doing so, we introduce the following conditions.

Assumption 2:
Let ϵ be a relatively small positive number and M be a positive constant. Suppose that the following conditions hold:

1. *For $\Omega(\epsilon) = \{(\theta, g) : \ \|(\theta, g) - (\theta_0, g_0)\|_2 \leq \epsilon\}$,*

$$(a) \quad \sup_{(\theta,g)\in\Omega(\epsilon)} E\left\|g^{(2)}(x_1'\theta)x_1x_1'\right\|^2 \leq M;$$

$$(b) \quad \sup_{(\theta,g)\in\Omega(\epsilon)} \left\|\frac{1}{N}\sum_{i=1}^N \left(g^{(1)}(x_i'\theta)\right)^2 x_ix_i' - E\left[\left(g^{(1)}(x_1'\theta)\right)^2 x_1x_1'\right]\right\| = o_P(1);$$

$$(c) \quad \sup_{(\theta_1,g_1),(\theta_2,g_2)\in\Omega(\epsilon)} \left\|\frac{1}{N}\sum_{i=1}^N g_1(x_i'\theta_1)g_2^{(2)}(x_i'\theta_2)x_ix_i' - E\left[g_1(x_1'\theta_1)g_2^{(2)}(x_1'\theta_2)x_1x_1'\right]\right\| = o_P(1).$$

2. *Let $\Sigma_1(\theta) = E[\dot{\mathcal{H}}(x_1'\theta)\dot{\mathcal{H}}(x_1'\theta)']$ and $\Sigma_2(\theta_0) = E\left[\dot{\mathcal{H}}(x_1'\theta_0)\dot{\mathcal{H}}(x_1'\theta_0)'\|x_1\|^2\right]$, in which $\dot{\mathcal{H}}(w) = \left(h_1^{(1)}(w), \ldots, h_{k-1}^{(1)}(w)\right)'$. Suppose that $\sup_{\{\theta:\ \|\theta-\theta_0\|\leq\epsilon\}} \lambda_{max}\left(\Sigma_1(\theta)\right) \leq M_1$ and $\lambda_{max}\left(\Sigma_2(\theta_0)\right) \leq M_2$ for some $0 < M_1, M_2 < \infty$.*

3. *Suppose that (i) $E\left[\delta_{0,k}^{(1)}(x_1'\theta_0)\right]^4 = o(1)$ and (ii) $N/k^r \to 0$.*

For Assumption 2.1.a, arguments similar to those given in Remark 1 apply. Assumptions 2.1.b and 2.1.c are the same as Assumption 2 of Yu and Ruppert (2002), and require the uniform convergence in a small neighbourhood of (θ_0, g_0). We can further decompose these two conditions by using Lemma A2 of Newey and Powell (2003), and prove the uniform convergence by following a procedure similar to those given for Theorem 2.1. However, it will result in a quite lengthy development. For simplicity, we use Assumptions 2.1.a and 2.1.b in this paper. Assumption 2.2 is in the same spirit as Assumption 2 in Newey (1997) and Assumption 3.iv in Su and Jin (2012) but the derivatives of Hermite functions are involved. As the derivative of the classical orthogonal sequence is still orthogonal, the condition follows immediately. See Condition A.2 in Belloni et al. (2015).

In Assumption 2.3, conditions $E\left[\delta_{0,k}^{(1)}(x_1'\theta_0)\right]^4 = o(1)$ and $N/k^r \to 0$ ensure not only both $\delta_{0,k}^{(1)}(w)$ and $\delta_{0,k}(w)$ are sufficiently small respectively, but also they can be smoothed out when we establish the asymptotic normality. This is the so-called under–smoothing condition in the literature. See Belloni et al. (2015) and Chang et al. (2015).

In order to establish our asymptotic theory for the estimator, the score function and hessian matrix (the first and the second derivatives of $\mathcal{W}_{N,\lambda}(\theta, C_k)$, respectively) are investigated first in Lemma A.2 below. Due to the constraint $\|\theta_0\| = 1$ in the optimization (2.10), there is a projection matrix $P_{\theta_0} = I_d - \theta_0\theta_0'$ involved after carefully organising the score function. Note that P_{θ_0} will project any vector into the orthogonal complement space of θ_0, denoted by $\theta_0^{\perp}$, and $\theta_0^{\perp}$ is $(d-1)$-dimensional. Simple algebra shows P_{θ_0} has eigenvalues $0, 1, \ldots, 1$, where 0 has eigenvector θ_0 and each eigenvalue 1 corresponds to an eigenvector orthogonal to θ_0, denoted by p_j, $j = 1, \cdots, d-1$. Put $P_1 = (p_1, \ldots, p_{d-1})$ of $d \times (d-1)$. It then leads to $P_{\theta_0} = P_1 P_1'$ by a spectral decomposition and $P_1'P_1 = I_{d-1}$ straight away.

Note that (θ_0, P_1) is an orthogonal matrix that forms a new coordinate system in $\mathbb{R}^d$, under which $\widetilde{\theta} - \theta_0$ is expressed as

$$\begin{aligned}\widetilde{\theta} - \theta_0 &= (P_1P_1' + \theta_0\theta_0')(\widetilde{\theta} - \theta_0) = P_1[P_1'(\widetilde{\theta} - \theta_0)] + \theta_0(\theta_0'\widetilde{\theta} - 1) \\ &= (p_1, \ldots, p_{d-1})(\alpha_{N,1}, \ldots, \alpha_{N,d-1})' + \alpha_{N,0}\theta_0 \\ &= \alpha_{N,1}p_1 + \cdots + \alpha_{N,d-1}p_{d-1} + \alpha_{N,0}\theta_0, \end{aligned} \tag{2.11}$$

where we denote $\alpha_{N,0} = \theta_0'\widetilde{\theta} - 1$ and $(\alpha_{N,1}, \ldots, \alpha_{N,d-1})' = P_1'(\widetilde{\theta} - \theta_0)$, viz., they are coordinates of $\widetilde{\theta} - \theta_0$ under the new system. We now establish the following results.

Theorem 2.2. *Let Assumptions 1 and 2.1-2.3 hold.*

1. $\theta_0'\left(\widetilde{\theta} - \theta_0\right) = -\frac{1}{2}\|\widetilde{\theta} - \theta_0\|^2$.

2. *In addition, let Assumption 2.4 hold. Then as $N \to \infty$,*

$$\sqrt{N}\left(P_1'(\widetilde{\theta} - \theta_0) - 2(P_1'VP_1)^{-1}P_1'S_N(\widetilde{g}_k)\right) \to_D N(0, \sigma_e^2(P_1'VP_1)^{-1}),$$

where $V = E\big[\{g_0^{(1)}(x_1'\theta_0)\}^2 x_1x_1'\big]$ *and* $S_N(\widetilde{g}_k) = \frac{1}{N}\sum_{i=1}^N [g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\theta_0)]\, \widetilde{g}_k^{(1)}(x_i'\theta_0)x_i$.

We need fairly detailed explanation on the theorem in the following two remarks.

Remark 2. *As can be seen from the proof, the first result of Theorem 2.2 is purely due to the constraint* $\|\widetilde{\theta}\| = 1$ *in the optimization* (2.10). *Note that* $\theta_0'\left(\widetilde{\theta} - \theta_0\right) = \widetilde{\theta}'\theta_0 - 1$. *In the new coordinate system of* $\mathbb{R}^d$, $\widetilde{\theta}'\theta_0$ *is the coordinate of* $\widetilde{\theta}$ *at the axis of* θ_0. *Clearly,* $|\widetilde{\theta}'\theta_0| \le \|\widetilde{\theta}\|\|\theta_0\| = 1$ *by Cauchy-Schwarz inequality and the equality holds as long as* $\widetilde{\theta} = \theta_0$. *This is why the value of* $\theta_0'\left(\widetilde{\theta} - \theta_0\right)$ *is always negative. Notice also that* $1 - \widetilde{\theta}'\theta_0$ *measures the metric from the coordinate to the surface of the unit ball. Therefore, the first result indicates that, along the direction* θ_0, $\widetilde{\theta}$ *converges with a quicker rate than that in all other directions orthogonal to* θ_0. *More precisely, taking orthogonal expansion* (2.11) *into account, we have* $\|\widetilde{\theta} - \theta_0\|^2 = \|P_1'(\widetilde{\theta} - \theta_0)\|^2 + (\theta_0'\widetilde{\theta} - 1)^2$, *which, by plugging the first result, yields*

$$1 - \theta_0'\widetilde{\theta} = \frac{1}{1 + \theta_0'\widetilde{\theta}}\|P_1'(\widetilde{\theta} - \theta_0)\|^2 = \frac{1}{2}\|P_1'(\widetilde{\theta} - \theta_0)\|^2(1 + o_P(1))$$

due to Theorem 2.1. It follows from the second result that $1-\theta_0'\widetilde{\theta}$ can have convergence rate $O_P(N^{-1})$ especially in the situation discussed in the next remark.

Actually, the essential reason that we only consider the normality of $P_1'(\widetilde{\theta}-\theta_0)$ is because both $\widetilde{\theta}$ and θ_0 belong to the unit ball of $\mathbb{R}^d$ due to $\|\widetilde{\theta}\|=\|\theta_0\|=1$ and the unit ball, as a subspace of $\mathbb{R}^d$, has dimension $d-1$. Without P_1 that transforms $\widetilde{\theta}-\theta_0$ into $\mathbb{R}^{d-1}$, it is impossible to establish the normality of $\widetilde{\theta}-\theta_0$.

This fact is also verified from the proof of the theorem. Indeed, as stated before, the score $\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta}$ contains the projection matrix P_{θ_0} such that the score belongs to $\theta_0^{\perp}$, a $(d-1)$-dimensional subspace. Hence, the corresponding covariance matrix must be singular. This means, in order to obtain the normality, the score also needs to be rotated by P_1 into $(d-1)$-vector.

If $\sqrt{N}P_1'S_N(\widetilde{g}_k)\to_P 0$, then we have $\sqrt{N}P_1'(\widetilde{\theta}-\theta_0)\to_D N(0,\sigma_e^2(P_1'VP_1)^{-1})$. In this case, our estimate of $P_1'\widetilde{\theta}$ is asymptotically efficient, because $\sigma_e^2(P_1'VP_1)^{-1}$ reaches the information lower bound in the semiparametric sense (c.f. Carroll et al. (1997); Xia (2006)). Condition (2.6) of Theorem 2 in Chen et al. (2003) shares a similar issue on imposing a restriction like $\sqrt{N}S_N(\widetilde{g}_k)\to_P 0$, and they argue that the verification of this type of condition is in some cases difficult, and is itself the subject of a long paper as in Newey (1994). Although it is easy to show that $S_N(\widetilde{g}_k)$ is exactly 0 uniformly in N when $\widetilde{g}_k=g_0$, we cannot further decompose $\sqrt{N}P_1'S_N(\widetilde{g}_k)$ without more restrictive requirements. Newey (1994) does provide some assumptions (i.e., "Linearization" condition of Assumption 5.1) to deal with terms like $\sqrt{N}P_1'S_N(\widetilde{g}_k)$, but our simulation suggests that (2.10) does not always outperform the closed-form estimator provided below, which is very likely due to the fact that the biased term $\sqrt{N}P_1'S_N(\widetilde{g}_k)$ does not converge to 0 sufficiently fast. Therefore, we state Theorem 2.2 as it stands. In what follows we give one more example where the bias term in the theorem is negligible too.

Remark 3. *We now consider a special case where $\sqrt{N}P_1'S_N(g)$ is indeed negligible on a neighbourhood of g_0, $\{g:\|(g-g_0)g^{(1)}\|_{L^2}\le\epsilon_N\}$, where ϵ_N is a sequence of positive numbers and $\epsilon_N\to 0$ as $N\to\infty$. Suppose that $E[x_1]=0$ and $E[x_1x_1']=I_d$. Hence $E[P_1'x_1x_1'\theta_0]=P_1'\theta_0=0$, which indicates that $P_1'x_1$ and $x_1'\theta_0$ are uncorrelated. If we further suppose that x_1 is normally distributed and all elements of x_1 are independent of each other, then uncorrelatedness implies independence here. Thus, for each given $g\in\{g:\|(g-g_0)g^{(1)}\|_{L^2}\le\epsilon_N\}$, we have*

$$
\begin{aligned}
E\left\|\sqrt{N}P_1'S_N(g)\right\|^2 &= \frac{1}{N}E\left\|\sum_{i=1}^N\left[g_0(x_i'\theta_0)-g(x_i'\theta_0)\right]g^{(1)}(x_i'\theta_0)P_1'x_i\right\|^2\\
&=\frac{1}{N}\sum_{i=1}^N E\left[\left(g_0(x_i'\theta_0)-g(x_i'\theta_0)\right)g^{(1)}(x_i'\theta_0)\right]^2 E\left\|P_1'x_i\right\|^2\\
&=O(1)E\left[\left(g_0(x_1'\theta_0)-g(x_1'\theta_0)\right)g^{(1)}(x_1'\theta_0)\right]^2\\
&=O(1)\int\left[(g_0(w)-g(w))g^{(1)}(w)\right]^2 f_{\theta_0}(w)dw\le O(\epsilon_N), \qquad (2.12)
\end{aligned}
$$

where $f_{\theta_0}(w)$ is a normal density function as $w=x_1'\theta_0$ is normal. Thus, we obtain that for each g in the given neighbourhood, $\sqrt{N}P_1'S_N(g)=o_P(1)$. By using the proof similar to Lemma A2 of Newey and Powell (2003), it is easy to see that $\sup_{g\in\{g:\|(g-g_0)g^{(1)}\|_{L^2}\le\epsilon_N\}}\left\|\sqrt{N}P_1'S_N(g)\right\|=o_P(1)$. Noticing that ϵ_N is a arbitrary positive number, we can use, for example, $\epsilon_N=\frac{1}{\ln N}$ or $\epsilon_N=\|\widetilde{g}_k-g_0\|_{L^2}^{\nu}$

with ν being a sufficiently small positive number, which are certainly much slower than $\|\widetilde{g}_k - g_0\|_{L^2}$ by Section 2.3 with $\widetilde{g}_k$ being defined in Theorem 2.1 already. In other words, $\widetilde{g}_k$ certainly falls in $\{g : \|(g-g_0)g^{(1)}\|_{L^2} \le \epsilon_N\}$ with probability one. Thus, under this very special circumstance, the bias term in Theorem 2.2 is removable.

Given that we have shown several cases where the bias term is $o_P(1)$, we have the following corollary.

Corollary 2.1. *Let Assumptions 1 and 2 hold.*

1. *In addition, suppose that in Theorem 2.2, the bias term $\sqrt{N}P_1'S_N(\widetilde{g}_k) \to_P 0$, then we have as $N \to \infty$,*
$$\sqrt{N}P_1'(\widetilde{\theta} - \theta_0) \to_D N(0, \sigma_e^2(P_1'VP_1)^{-1}), \tag{2.13}$$
where $V = E\big[\{g_0^{(1)}(x_1'\theta_0)\}^2 x_1x_1'\big]$.
2. *Consequently, as $N \to \infty$, we have $\theta_0'\left(\widetilde{\theta} - \theta_0\right) = O_P\left(\frac{1}{N}\right)$.*

The proof of the corollary is omitted as it is clear from Theorem 2.2 and Remark 2. We will then move on to study the decomposition (2.5) in the next subsection.

2.2 Closed-Form Estimation Method

In order to establish a consistent closed-form estimate for θ_0 through (2.5), we explore the idea of double series expansion allured in Dong et al. (2015) for the general single-index modelling of cross-sectional data. We first define an ordering relationship with respect to p in (2.5).

Definition 2.1. *Let $P_m = \{\mathsf{p} : |\mathsf{p}| = m\}$, where m is a non-negative integer. Suppose that $\hat{\mathsf{p}}, \check{\mathsf{p}} \in P_m$. We say $\hat{\mathsf{p}} = (\hat{\mathsf{p}}_1, \ldots, \hat{\mathsf{p}}_d) < \check{\mathsf{p}} = (\check{\mathsf{p}}_1, \ldots, \check{\mathsf{p}}_d)$ if $\hat{\mathsf{p}}_j = \check{\mathsf{p}}_j$ for all $j = 1, \ldots, l-1$ and $\hat{\mathsf{p}}_l < \check{\mathsf{p}}_l$, where $1 < l \le d$.*

By Definition 2.1, we list all $\mathscr{H}_{\mathsf{p}}(x_i)$'s in descending order with respect to $P_m = \{\mathsf{p} : |\mathsf{p}| = m\}$ for $m = 0, 1, \cdots, k-1$ below.

- As $m = 0$, p in (2.5) has only one realization $\mathsf{p} = (0, 0, \ldots, 0)'$. Thus,
$$\mathscr{H}_{\mathsf{p}}(x_i) \equiv 1 \quad \text{and} \quad a_{0,\mathsf{p}}(\theta_0) = c_{0,0}. \tag{2.14}$$

- As $m = 1$, p in (2.5) has d possibilities:
$$\begin{aligned} &\mathsf{p} = (1, 0, \ldots, 0)',\ \mathscr{H}_{\mathsf{p}}(x_i) = x_{i,1},\ a_{1,\mathsf{p}}(\theta_0) = c_{0,1}\theta_{0,1}, \\ &\quad\vdots \\ &\mathsf{p} = (0, \ldots, 0, 1)',\ \mathscr{H}_{\mathsf{p}}(x_i) = x_{i,d},\ a_{1,\mathsf{p}}(\theta_0) = c_{0,1}\theta_{0,d}. \end{aligned} \tag{2.15}$$

- As $m = 2, \ldots, k-1$, we have $m^* := \binom{m+d-1}{d-1}$ possibilities of vector p:
$$\begin{aligned} &\mathsf{p} = (m, 0, \ldots, 0)',\ \mathscr{H}_{\mathsf{p}}(x_i) = h_m(x_{i,1}),\ a_{m,\mathsf{p}}(\theta_0) = c_{0,m}\theta_{0,1}^m, \\ &\mathsf{p} = (m-1, 1, 0, \ldots, 0)',\ \mathscr{H}_{\mathsf{p}}(x_i) = h_{m-1}(x_{i,1})\,h_1(x_{i,2}),\ a_{m,\mathsf{p}}(\theta_0) = \sqrt{m}c_{0,m}\theta_{0,1}^{m-1}\theta_{0,2}, \end{aligned}$$

$$\vdots$$

$$\mathsf{p} = (0,\ldots,0,m)',\ \mathscr{H}_{\mathsf{p}}(x_i) = h_m(x_{i,d}),\ a_{m,\mathsf{p}}(\theta_0) = c_{0,m}\theta_{0,d}^m. \tag{2.16}$$

Therefore, a detailed ordering for all p's in (2.14)-(2.16) is as follows:

$$\begin{aligned}
P_0: &\quad \mathsf{p}_1 = (0,0,\ldots,0),\\
P_1: &\quad \begin{cases} \mathsf{p}_2 = (1,0,\ldots,0),\\ \vdots \\ \mathsf{p}_{d+1} = (0,0,\ldots,1), \end{cases}\\
&\vdots\\
P_{k-1}: &\quad \begin{cases} \mathsf{p}_{\frac{(d+k-2)!}{d!(k-2)!}+1} = (k-1,0,\ldots,0),\\ \mathsf{p}_{\frac{(d+k-2)!}{d!(k-2)!}+2} = (k-2,1,\ldots,0),\\ \vdots \\ \mathsf{p}_{\frac{(d+k-1)!}{d!(k-1)!}} = (0,0,\ldots,k-1). \end{cases}
\end{aligned} \tag{2.17}$$

Then it allows us to rewrite model (1.1) as

$$y_i = \mathcal{Z}(x_i)'\beta_{0,K} + \delta_{0,k}(x_i'\theta_0) + e_i, \tag{2.18}$$

where $\mathcal{Z}(x_i) = (z_1(x_i),\ldots,z_K(x_i))'$, each $z_j(x_i)$ for $j = 1,\ldots,K$ represents $\mathscr{H}_{\mathsf{p}}(x_i)$ associated with p based on (2.14)-(2.17), and $\beta_{0,K}$ is a $K\times 1$ vector consisting of the corresponding coefficients $a_{m,\mathsf{p}}(\theta_0)$'s. Moreover, it is easy to check that the cardinality of $P_m = \{\mathsf{p} : |\mathsf{p}| = m\}$ in Definition 2.1 is $\binom{m+d-1}{d-1}$ for $m = 0,1\ldots,k-1$, so we can calculate the length of $\mathcal{Z}(w)$ as $K = \sum_{m=0}^{k-1}\binom{m+d-1}{d-1} = \frac{(d+k-1)!}{d!(k-1)!} = O(k^d)$.

Notice that in equation (2.18), the regressors $\{x_i, i\geq 1\}$ contained in $\{\mathcal{Z}(x_i), i\geq 1\}$ and the unknown parameter $\beta_{0,K}$ are separated completely and form an approximate linear model. Since all the functions in the vector $\mathcal{Z}(\cdot)$ are known, model (2.18) is actually parametrically linear. Hence, the OLS method gives an estimate of the form:

$$\widehat{\beta}_K = \left(\sum_{i=1}^N \mathcal{Z}(x_i)\mathcal{Z}(x_i)'\right)^{-1}\sum_{i=1}^N \mathcal{Z}(x_i)y_i. \tag{2.19}$$

In order to investigate (2.19), we first introduce the following conditions.

Assumption 3: *Let $f(\cdot)$ be the pdf of x_1.*

1. *Suppose* $\exp\left(\|x\|^2/2\right)f(x)\leq M$ *uniformly on $\mathbb{R}^d$.*
2. *Suppose $0<\rho_1\leq\lambda_{min}(\mathcal{Z}_x)$ uniformly in k, where $\lambda_{min}(\mathcal{Z}_x)$ denotes the minimum eigenvalue of $\mathcal{Z}_x$, $\mathcal{Z}_x = E\left[\mathcal{Z}(x_1)\mathcal{Z}(x_1)'\right]$, and ρ_1 is a positive constant.*
3. *Let $\Psi(N,k) = \frac{1}{N}\sum_{m=1}^K\sum_{n=1}^K\int_{\mathbb{R}^d}|\mathscr{H}_{\mathsf{p}_m}(z)|^2|\mathscr{H}_{\mathsf{p}_n}(z)|^2 f(z)dz$, where p_m and p_n are defined in (2.17) for $m,n = 1,\ldots,K$. As $(N,k)\to(\infty,\infty)$, suppose $\Psi(N,k)\to 0$.*

Assumption 3.1 imposes some restriction on the pdf of x, and excludes heavy-tailed distributions. Assumption 3.2 is standard in the literature (e.g., Assumption 2 of Newey (1997) and Assumption 3.iii of Su and Jin (2012)). Again, in this paper, we consider the function space $L^2(\mathbb{R}, \exp(-w^2/2))$ rather than $L^2(\mathbb{R})$, thereby we use Hermite polynomials instead of Hermite functions to decompose the link function. It entails that $\mathscr{H}_{\mathsf{p}_m}(z)$'s are not uniformly bounded as those used in Dong et al. (2016), which is the reason why we introduce Assumption 3.3 herewith.

Remark 4. *For $\Psi(N,k)$, there are two ways to simplify the notation:*

1. *Impose a stronger version of Assumption 3.1 as follows:*
   - *Suppose* $\exp(\|x\|^2) f(z) \leq M$ *uniformly on* $\mathbb{R}^d$.

   Then, we are able to further organise $\Psi(N,k)$:

$$\begin{aligned}
\Psi(N,k) &= \frac{1}{N}\sum_{m=1}^{K}\sum_{n=1}^{K}\int_{\mathbb{R}^d} |\mathscr{H}_{\mathsf{p}_m}(x)|^2 |\mathscr{H}_{\mathsf{p}_n}(x)|^2 \exp(-\|x\|^2)\cdot\exp(\|x\|^2) f(x)dx \\
&\leq O(1)\frac{1}{N}\sum_{m=1}^{K}\sum_{n=1}^{K}\int_{\mathbb{R}^d} |\mathscr{H}_{\mathsf{p}_m}(x)|^2 |\mathscr{H}_{\mathsf{p}_n}(x)|^2 \exp(-\|x\|^2)dx \\
&\leq O(1)\frac{K}{N}\sum_{m=1}^{K}\int_{\mathbb{R}^d} |\mathscr{H}_{\mathsf{p}_m}(x)|^2 \exp(-\|x\|^2/2)dx = O\left(\frac{K^2}{N}\right) = O\left(\frac{k^{2d}}{N}\right).
\end{aligned}$$

2. *We can follow Assumption 3.iii of Su and Jin (2012), i.e., assuming $E|\mathscr{H}_{\mathsf{p}}(x)|^4$ is uniformly bounded.*

Either restriction above will make $\Psi(N,k)$ reduce to $\frac{k^{2d}}{N}$ under the restrictive condition on the pdf of x_1. Indeed, as the link function $g_0(w)$ is potentially non-integrable and unbounded, one may need to impose some restrictions on the pdf of the regressors.

Lemma 2.1. *Under Assumptions 1.1, 1.3 and 3, as $N \to \infty$,*

$$\|\widehat{\beta}_K - \beta_{0,K}\| = O_P\left(\sqrt{\frac{k^d}{N}}\right) + O_P(k^{-r/2}).$$

Notice that the leading term of the rate in Lemma 2.1 is $k^{d/2}N^{-1/2}$, which indicates the curse of dimensionality caused by the decomposition (2.5), and thus is a price we have to pay for establishing a consistent closed-form estimate for θ_0. On the other hand, it allows us to drop the assumption "Θ is a compact set" and avoid being bothered by the residual term $V^{-1}S_N(\widetilde{g}_k)$ involved in Theorem 2.2.

We are ready to proceed further to establish an asymptotically consistent estimate for θ_0.

A case of $c_{0,1} \neq 0$**.** In view of (2.15), the 2^{nd} to $(d+1)^{th}$ elements of $\widehat{\beta}_K$ are the estimates of $c_{0,1}\theta_{0,1}, \ldots, c_{0,1}\theta_{0,d}$. Moreover, because of Lemma 2.1 and $\|\widehat{\beta}_K^d - \beta_{0,K}^d\| \leq \|\widehat{\beta}_K - \beta_{0,K}\|$, where $\widehat{\beta}_K^d$ and $\beta_{0,K}^d$ denote vectors of the 2^{nd} to $(d+1)^{th}$ elements of $\widehat{\beta}_K$ and $\beta_{0,K}$, respectively, the estimates of $c_{0,1}\theta_{0,1}, \ldots, c_{0,1}\theta_{0,d}$ are consistent. Therefore, it follows from the identification condition and the continuous mapping theorem that $\|\widehat{\beta}_K^d\| = \left(\sum_{j=2}^{d+1}\widehat{\beta}_{K,j}^2\right)^{1/2} \to_P |c_{0,1}|$, where $\widehat{\beta}_{K,j}$ denotes the j^{th}

element of $\widehat{\beta}_K$. Since $\theta_{0,1} > 0$, $\widehat{\beta}_{K,2} \neq 0$ and $\widehat{\beta}_{K,2}$ has the same sign as $c_{0,1}$ with a probability approaching one, we define the estimate of θ_0 as

$$\widehat{\theta} = \frac{1}{\widehat{c}_1} Q\widehat{\beta}_K, \tag{2.20}$$

where $\widehat{c}_1 = \text{sgn}(\widehat{\beta}_{K,2})\|\widehat{\beta}_K^d\|$ and $Q = \left(0_{d\times 1},\ I_d,\ 0_{d\times(K-d-1)}\right)$. With the above set-ups, the following theorem establishes an asymptotic normality for $\widehat{\theta}$ given in (2.20).

Theorem 2.3. *Under Assumptions 1.1, 1.3 and 2.4.ii and 3. As $N \to \infty$,*

$$\sqrt{N} P_1' \left(\widehat{\theta} - \theta_0\right) \to_D N\left(0,\ \sigma_e^2 c_{0,1}^{-2}\Omega_1\right),$$

where $\Omega_1 = \lim_{k\to\infty} P_1' Q \mathcal{Z}_x^{-1} Q' P_1$, and $\mathcal{Z}_x = E[\mathcal{Z}(x_1)\mathcal{Z}(x_1)']$ is defined in Assumption 3.

According to Theorem 2.3, it is obvious that for the closed-form estimate (2.20), the asymptotic efficiency is heavily affected by $c_{0,1}$. We will further confirm this argument in the Monte Carlo study later on.

Remark 5. *We now take a further look at the asymptotic covariance matrices of Theorems 2.2 and 2.3 in order to compare the efficiency. Since they both share the same term σ_e^2, we will ignore it in the following discussion. For the asymptotic covariance matrix given in Theorem 2.3 (i.e., $c_{0,1}^{-2}\Omega_1$), we now reshuffle $\mathcal{Z}(x_1)$ a little bit and let the first d elements be $x_1 = (x_{1,1}, \ldots, x_{1,d})'$. This can easily be achievable, because the order of $\mathcal{Z}(x_1)$ is totally subjective. Then we need only to consider $c_{0,1}^{-2} P_1' \widetilde{Q} \mathcal{Z}_x^{-1} \widetilde{Q}' P_1$, where $\widetilde{Q} = \left(I_d,\ 0_{d\times(K-d)}\right)$. Partition $\mathcal{Z}_x = E[\mathcal{Z}(x_1)\mathcal{Z}(x_1)']$ as*

$$\mathcal{Z}_x = \begin{pmatrix} \mathcal{Z}_{x,11} & \mathcal{Z}_{x,12} \\ \mathcal{Z}_{x,21} & \mathcal{Z}_{x,22} \end{pmatrix},$$

where $\mathcal{Z}_{x,11}$ is the $d \times d$ principal sub-matrix of $\mathcal{Z}_x$. Then $\mathcal{Z}_x^{-1}$ can be expressed as

$$\mathcal{Z}_x^{-1} = \begin{pmatrix} \widetilde{\mathcal{Z}}_{x,11}^{-1} & \widetilde{\mathcal{Z}}_{x,11}^{-1}\mathcal{Z}_{x,12}\mathcal{Z}_{x,22}^{-1} \\ \mathcal{Z}_{x,22}^{-1}\mathcal{Z}_{x,21}\widetilde{\mathcal{Z}}_{x,11}^{-1} & \mathcal{Z}_{x,22}^{-1} + \mathcal{Z}_{x,22}^{-1}\mathcal{Z}_{x,21}\mathcal{Z}_{x,11}^{-1}\mathcal{Z}_{x,12}\mathcal{Z}_{x,22}^{-1} \end{pmatrix},$$

where $\widetilde{\mathcal{Z}}_{x,11} = \mathcal{Z}_{x,11} - \mathcal{Z}_{x,12}\mathcal{Z}_{x,22}^{-1}\mathcal{Z}_{x,21}$ and $\mathcal{Z}_{x,11} = E[x_1 x_1']$. Thus, we are able to write

$$c_{0,1}^{-2}\widetilde{Q}\mathcal{Z}_x^{-1}\widetilde{Q}' = c_{0,1}^{-2}\widetilde{\mathcal{Z}}_{x,11}^{-1} = c_{0,1}^{-2}\left(\mathcal{Z}_{x,11} - \mathcal{Z}_{x,12}\mathcal{Z}_{x,22}^{-1}\mathcal{Z}_{x,21}\right)^{-1}. \tag{2.21}$$

Based on the above partition, it is clear that the difference between two asymptotic covariance matrices lies at the difference between

$$(P_1' V P_1)^{-1} \quad \text{and} \quad c_{0,1}^{-2} P_1' \left(\mathcal{Z}_{x,11} - \mathcal{Z}_{x,12}\mathcal{Z}_{x,22}^{-1}\mathcal{Z}_{x,21}\right)^{-1} P_1. \tag{2.22}$$

In general, it is hard to compare them, so we focus on some special cases below.

Further suppose that the elements, $x_{1,1}, \cdots, x_{1,d}$, of $x_1 = (x_{1,1}, \ldots, x_{1,d})'$ are independent of each other and the density of each element $x_{1,j}$ for $j = 1, \ldots, d$ is $f_j(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$. Then it is easy *to check that $\mathcal{Z}_{x,12}$ and $\mathcal{Z}_{x,12}$ become 0 matrices, so the comparison (2.22) immediately reduces to*

$$(P_1'VP_1)^{-1} = \left\{P_1'E\Big[\left(g_0^{(1)}(\theta_0'x_1)\right)^2 x_1x_1'\Big]P_1\right\}^{-1} \quad \text{and} \quad c_{0,1}^{-2}\, I_{d-1}. \tag{2.23}$$

Therefore, it is clear that if $|g_0^{(1)}(w)|$ is bounded away from zero and above from infinity, two estimates have equivalent efficiency up to a constant, but this constant depends on the true regression function only because of $c_{0,1} = g_0^{(1)}(0)$. In particular, for the linear models $g_0(w) = c_{0,1}w + c_{0,0}$, we can see $(P_1'VP_1)^{-1} = (c_{0,1}^2 P_1'E[xx']P_1)^{-1} = c_{0,1}^{-2}\, I_{d-1}$, which means both estimates are exactly the same in terms of the forms of asymptotic covariance matrices.

We now focus on testing $c_{0,1}$ (i.e., $H_0:\ c_{0,1} = 0;\ H_1:\ c_{0,1} \neq 0$), which is equivalent to testing the joint significance of $Q\beta_{0,K}$ due to the fact that $\|\theta_0\| = 1$. It is easy to show that

$$\widehat{\sigma}_e^2 = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \mathcal{Z}(x_i)'\widehat{\beta}_K\right)^2 \to_P \sigma_e^2,$$

$$\widehat{\Xi} = \widehat{\sigma}_e^2 Q\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)\mathcal{Z}(x_i)'\right)^{-1} Q' \to_P \sigma_e^2\,\Omega_1, \tag{2.24}$$

which gives $\widehat{\Xi}^{-1/2}\cdot\sqrt{N}\left(Q\widehat{\beta}_K - Q\beta_{0,K}\right) \to_D N(0, I_d)$. The significance test can be established by the following corollary.

Corollary 2.2. *Let the conditions of Theorem 2.3 hold. Under the null (i.e., $H_0:\ c_{0,1} = 0$), $N\widehat{\beta}_K'Q'\widehat{\Xi}^{-1}Q\widehat{\beta}_K \to_D \chi^2(d)$, where $\widehat{\Xi}$ is defined in (2.24).*

A general case that $c_{0,m_0} \neq 0$ with $m_0 \geq 1$ an integer. Such m_0 does exist, otherwise g_0 becomes a constant function. Note that the coefficients for $h_{m_0}(x_{i,1}), \ldots, h_{m_0}(x_{i,d})$ are $c_{0,m_0}\theta_{0,1}^{m_0}, \ldots, c_{0,m_0}\theta_{0,d}^{m_0}$ by (2.16). The corresponding estimates of these coefficients are the $m_1^{th}, \ldots, m_d^{th}$ elements of $\widehat{\beta}_K$, respectively, where $m_1 = \frac{(d+m_0-1)!}{d!(m_0-1)!} + 1$, $m_2 = m_1 + \frac{(d+m_0-1)!}{(d-1)!m_0!} - \frac{(d+m_0-2)!}{(d-2)!m_0!} + 1$, $m_3 = m_2 + \frac{(d+m_0-2)!}{(d-2)!m_0!} - \frac{(d+m_0-3)!}{(d-3)!m_0!} + 1$, $\ldots$, $m_d = \frac{(d+m_0)!}{d!m_0!}$. Since m_0 is fixed, it follows from Lemma 2.1 and the continuous mapping theorem that $\left(\sum_{i=1}^d(\widehat{\beta}_{K,m_i}^2)^{1/m_0}\right)^{m_0/2} \to_P |c_{m_0}|$ by the identification restriction imposed on θ_0, where $\widehat{\beta}_{K,m_j}$ denotes the m_j^{th} element of $\widehat{\beta}_K$ for $j = 1, \ldots, d$. Hence, by the same reason as the case of $c_{0,1} \neq 0$, a consistent estimate for c_{0,m_0} is $\widehat{c}_{m_0} = \text{sgn}(\widehat{\beta}_{K,m_1})\left(\sum_{i=1}^d(\widehat{\beta}_{K,m_i}^2)^{1/m_0}\right)^{m_0/2}$.

Meanwhile, because again $\widehat{\beta}_{K,m_1}$ is the consistent estimate of $c_{0,m_0}\theta_{0,1}^{m_0}$, it follows from the continuous mapping theorem that $\widehat{\theta}_{0,1} = (\widehat{\beta}_{K,m_1}/\widehat{c}_{m_0})^{1/m_0}$ is a consistent estimate for $\theta_{0,1}$. Let us construct Q in the same fashion as before such that $Q\widehat{\beta}_K$ consists of the m_1^{th}, $(m_1+1)^{th}$, $\cdots$, $(m_1+d-1)^{th}$ elements in $\widehat{\beta}_K$. These elements are the estimates of $c_{0,m_0}\theta_{0,1}^{m_0}$, $\sqrt{m_0}c_{0,m_0}\theta_{0,1}^{m_0-1}\theta_{0,2}$, $\cdots$, $\sqrt{m_0}c_{0,m_0}\theta_{0,1}^{m_0-1}\theta_{0,d}$, respectively. With $\widehat{c}_{m_0}$ and $\widehat{\theta}_{0,1}$ at hand, we are finally able to define an estimate for θ_0 of the form:

$$\widehat{\theta} = \widehat{Q}_1\, Q\, \widehat{\beta}_K \quad \text{and} \quad \widehat{Q}_1 = \text{diag}\left(\frac{1}{\widehat{c}_{m_0}\widehat{\theta}_{0,1}^{m_0-1}}, \frac{\widehat{\theta}_{0,1}}{\sqrt{m_0}\widehat{\beta}_{m_1}}, \cdots, \frac{\widehat{\theta}_{0,1}}{\sqrt{m_0}\widehat{\beta}_{m_1}}\right). \tag{2.25}$$

The following theorem establishes another asymptotic normality for $\widehat{\theta}$ of (2.25).

Theorem 2.4. *Suppose that* $m_0 \geq 1$ *is an integer such that* $c_{0,m_0} \neq 0$. *Under the conditions of Theorem 2.3, as* $N \to \infty$, *the estimate of* (2.25) *obeys*

$$\sqrt{N} P_1' \left(\widehat{\theta} - \theta_0 \right) \to_D N(0, \sigma_e^2 \Omega_2),$$

where $Q_1 = c_{0,m_0}^{-1} \theta_{0,1}^{-m_o+1} diag(1, m_0^{-1/2}, \ldots, m_0^{-1/2})$, $\Omega_2 = \lim_{k\to\infty} P_1' Q_1 Q \mathcal{Z}_x^{-1} Q' Q_1 P_1$, *and* $\mathcal{Z}_x$ *is denoted in Assumption 3.*

Again, for the closed-form estimate given by (2.25), the asymptotic efficiency is heavily affected by c_{0,m_0}. Also, we can compare the efficiency for Theorem 2.2 and Theorem 2.4 for some special cases as in Remark 5.

Following the spirit of Corollary 2.2, in order to test if $c_{0,m_0} = 0$, we need only to implement the following test.

Corollary 2.3. *Let the conditions of Theorem 2.4 hold. For a fixed positive integer* $m_0 \geq 1$, *under the null (i.e.,* $H_0 : \ c_{0,m_0} = 0$*),* $N \widehat{\beta}_K' Q' \widehat{\Xi}^{-1} Q \widehat{\beta}_K \to_D \chi^2(d)$, *where* $\widehat{\Xi}$ *is defined in (2.24) and* Q *is denoted in (2.25).*

Notice that Theorem 2.4 and Corollary 2.3 simply reduce to Theorem 2.3 and Corollary 2.2, respectively, when $m_0 = 1$. When we need to implement the closed-form estimate, how to choose m_0 remains unknown to us. One fact is that smaller m_0 involves less computation, which may lead to more accurate approximation in practice.

2.3 Estimation of the link function

Based on the two methods provided in Sections 2.1 and 2.2, we now assume that we already obtain a consistent estimate, still denoted by $\widehat{\theta}$, which satisfies $\widehat{\theta} - \theta_0 = O_P\left(\frac{1}{\sqrt{N}}\right)$. To recover the unknown link function, $\widehat{\theta}$ is plugged in (2.9) and then the estimator of $C_{0,k}$ is obtained as

$$\widehat{C}_k = \left[\sum_{i=1}^{N} \mathcal{H}(x_i'\widehat{\theta}) \mathcal{H}(x_i'\widehat{\theta})' \right]^{-1} \sum_{i=1}^{N} \mathcal{H}(x_i'\widehat{\theta}) y_i. \tag{2.26}$$

The following assumption is necessary to facilitate the development.

Assumption 4: *Let* ϵ *be a relatively small positive number. As* $(N, k) \to (\infty, \infty)$, *let* $\Phi(N, k) \to 0$ *uniformly in* $\theta \in \{\theta : \|\theta - \theta_0\| < \epsilon\}$, *where*

$$\Phi(N, k) = \frac{1}{N} \sum_{m=1}^{k} \sum_{n=1}^{k} \int_{\mathbb{R}} h_m^2(z) h_n^2(z) f_\theta(z) dz,$$

where $f_\theta(z)$ *is the pdf of* $z = x'\theta$ *as defined in Assumption 1.*

Assumption 4 is in the same spirit as Assumption 3.3. The discussion similar to those given in Remark 4 applies here. Further investigating (2.26) with Assumption 4, we then obtain the following theorem.

Theorem 2.5. *Assume that* $\widehat{\theta} - \theta_0 = O_P\left(\frac{1}{\sqrt{N}}\right)$. *Under Assumptions 1, 2.3 and 4, as* $N \to \infty$,

1. $\|\widehat{C}_k - C_{0,k}\| = O_P\left(\sqrt{\frac{k}{N}}\right) + O_P\left(k^{-r/2}\right)$, *where* $\widehat{C}_k$ *is denoted in (2.26);*

2. $\|\widehat{g}_k(w) - g_0(w)\|_{L^2} = O_P\left(\sqrt{\frac{k}{N}}\right) + O_P\left(k^{-r/2}\right)$, *where* $\widehat{g}_k(w) = \widehat{C}_k'\mathcal{H}(w)$.

So far, we have successfully established a consistent closed-form estimates for θ_0 and then the link function g_0. In the next section, we will evaluate the proposed model and the estimation methods using an extensive Monte Carlo study.

3 Monte Carlo Study

In this section, we perform a Monte Carlo study to investigate the finite sample properties of our estimates based on 10,000 replications. The data generating process (DGP) for model (1.1) is as follows:

Let $\theta_0 = (\theta_{0,1}, \theta_{0,2})' = (0.8, -0.6)'$. For regressor x_i, suppose that $x_i = (x_{i,1}, x_{i,2})'$ with $x_{i,1} \sim N(1,1)$ and $x_{i,2} \sim N(0,4)$. For the error term, assume $e_i \sim N(0,1)$. In order to examine the arguments that we made under Theorems 2.3 and 2.4, we consider the next 4 options for $g_0(w)$:

$$\begin{aligned}
&(a).\ g_0(w) = 0.5(h_1(w) + h_3(w)) = 0.5(w^3 - 2w),\\
&(b).\ g_0(w) = 10(h_1(w) + h_3(w)) = 10(w^3 - 2w),\\
&(c).\ g_0(w) = 0.5\exp(w),\\
&(d).\ g_0(w) = 10\exp(w).
\end{aligned} \tag{3.1}$$

Throughout the whole simulation process, the truncation parameter is always chosen as $k = \lfloor N^{1/4} \rfloor$ (see the Monte Carlo studies of Su and Jin (2012) and Dong et al. (2016) for similar setting). Although $k = \lfloor N^{1/4} \rfloor$ may not be an optimal choice, the asymptotic results above remain valid. For the case of nonparametric models, while existing studies have examined the choice of optimal truncation parameter (e.g., Gao et al. (2002)), it does not seem that there is such work available about how to choose an optimal k for the single-index model to the best of our knowledge. Thus we leave it for future work.

3.1 Hypothesis Test

In order to investigate the finite sample properties of the hypothesis test given in Corollary 2.3, we implement the following three hypotheses for each generated data set:

$$\begin{aligned}
&\text{Hypothesis 1}: \text{“}H_0: c_{0,1} = 0,\ H_1: c_{0,1} \neq 0\text{”};\\
&\text{Hypothesis 2}: \text{“}H_0: c_{0,2} = 0,\ H_1: c_{0,2} \neq 0\text{”};\\
&\text{Hypothesis 3}: \text{“}H_0: c_{0,3} = 0,\ H_1: c_{0,3} \neq 0\text{”}.
\end{aligned}$$

Specifically, for each replication, we calculate the test statistic by Corollary 2.3 and compare it with the critical value (at 5% significant level) with respect to the distribution of $\chi^2(2)$. After 10,000 replications, we then report the probability that we reject null. According to the four options defined in (3.1), we expect to reject both "$H_0: c_{0,1} = 0$" and "$H_0: c_{0,3} = 0$" for all $(a) - (d)$ (i.e., testing power).

Also, we expect to accept "$H_0: \ c_{0,2} = 0$" for (a) and (b) (i.e., testing size) and reject "$H_0: \ c_{0,2} = 0$" for (c) and (d) (i.e., testing power). The results are reported below in Table 1.

Table 1: The Size and Power of Hypothesis Tests

		Null Hypothesis		
$g_0(w)$ in (3.1)	N	$c_{0,1} = 0$	$c_{0,2} = 0$	$c_{0,3} = 0$
(a)	400	1.0000	0.0556	1.0000
	800	1.0000	0.0572	1.0000
	1600	1.0000	0.0509	1.0000
(b)	400	1.0000	0.0563	1.0000
	800	1.0000	0.0571	1.0000
	1600	1.0000	0.0510	1.0000
(c)	400	0.9996	0.9956	0.9998
	800	1.0000	1.0000	0.9964
	1600	1.0000	1.0000	1.0000
(d)	400	0.9997	0.9974	0.9999
	800	1.0000	1.0000	0.9998
	1600	1.0000	1.0000	1.0000

[1]Cells in the box indicate size.

As we can see from Table 1, the proposed test (i.e., Corollary 2.3) has good size and power values, which are close to what we expect at 5% significant level.

3.2 Bias and RMSE

For each option of $g_0(w)$ given in (3.1), we estimate θ_0 by $\widehat{\theta}$, denoted for either $\widetilde{\theta}$ in (2.10) or $\widehat{\theta}$ in (2.25) for notational simplicity.[3] We consider the case of $m_0 = 1, 2, 3$. Moreover, for each generated data set and each estimate $\widehat{\theta} = (\widehat{\theta}_1, \widehat{\theta}_2)'$, we record the bias and squared bias for $\widehat{\theta}_j$ with $j = 1, 2$. After 10,000 replications, we report the mean bias and root mean squared error (RMSE) results in Table 2 below. Note that compared to (b) and (d) in (3.1), options (a) and (c) have relatively small $c_{0,1}$ and $c_{0,3}$. Therefore, if we can choose m_0 correctly, we expect the closed-form estimates to have better performances for options (b) and (d). Specifically, for $m_0 = 1, 3$, we expect the closed-form estimators to have smaller bias and RMSE for options (b) and (d); although we fail to reject "$H_0: \ c_{0,2} = 2$" at a quite high probability for options (a) and (b), we still report the bias and RMSE for these two cases as a comparison.

As can be seen from Table 2, the estimates in the dark cells have relatively large bias and RMSE values. Apart from the dark cells, RMSEs generally decrease as the sample size increases. For the closed-form estimates, the overall RMSEs for options (b) and (d) are much smaller than those of options (a) and (c). Although we have pointed out that the asymptotic covariance of Theorem 2.2 reaches the information lower bound, it seems that (2.10) does not out-perform (2.25) in general. One possible reason is due to the bias caused by the term $(P_1'VP_1)^{-1}P_1'S_N(\widetilde{g}_k)$ in Theorem 2.2.

[3]By (1) of Theorem 2.1, we can explicitly calculate the fast rate of $1 - \widetilde{\theta}'\theta_0$ for the method studied in Section 2.1, because θ_0 is known and $\widetilde{\theta}$ is achievable for our simulation study. As the value of $1 - \widetilde{\theta}'\theta_0$ is already reflected by the biases and RMSEs reported in Table 2, we do not calculate it again for the purpose of conciseness.

Table 2: Bias and MSE

	Estimator		(2.10)		(2.25)					
					$m_o = 1$		$m_o = 2$		$m_o = 3$	
	$g_o(w)$ in (3.1)	N	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$	$\widehat{\theta}_1$	$\widehat{\theta}_2$
Bias	(a)	400	-0.0008	0.0001	-0.0065	-0.0010	0.0024	0.6012	-0.0036	-0.0022
		800	-0.0006	-0.0003	-0.0032	-0.0001	0.0001	0.5865	-0.0028	-0.0014
		1600	-0.0005	-0.0001	-0.0019	-0.0004	0.0005	0.6012	-0.0017	-0.0013
	(b)	400	-0.0004	0.0012	0.0000	0.0000	0.0023	0.6029	0.0000	0.0000
		800	-0.0008	-0.0006	0.0000	0.0000	-0.0006	0.5864	0.0000	0.0000
		1600	-0.0006	-0.0001	0.0000	0.0000	0.0010	0.6018	0.0000	0.0000
	(c)	400	-0.0019	-0.0004	-0.0031	0.0003	-0.0149	-0.0005	-0.0353	-0.0154
		800	-0.0013	-0.0005	-0.0012	0.0000	-0.0056	-0.0021	-0.0421	-0.0116
		1600	-0.0006	-0.0001	-0.0007	-0.0001	-0.0030	-0.0013	-0.0208	-0.0095
	(d)	400	-0.0013	0.0004	-0.0003	0.0008	-0.0043	0.0027	-0.0066	-0.0036
		800	-0.0010	0.0001	-0.0001	-0.0001	-0.0002	0.0001	-0.0049	-0.0002
		1600	-0.0009	0.0000	0.0001	0.0001	0.0000	0.0000	0.0002	0.0007
RMSE	(a)	400	0.0238	0.0279	0.0593	0.0753	0.2580	0.8068	0.0345	0.0442
		800	0.0168	0.0199	0.0437	0.0567	0.2625	0.7967	0.0327	0.0420
		1600	0.0158	0.0190	0.0306	0.0401	0.2571	0.8090	0.0214	0.0279
	(b)	400	0.0243	0.0374	0.0028	0.0038	0.2578	0.8083	0.0017	0.0022
		800	0.0159	0.0204	0.0021	0.0028	0.2626	0.7973	0.0016	0.0021
		1600	0.0178	0.0232	0.0015	0.0020	0.2566	0.8091	0.0010	0.0014
	(c)	400	0.0312	0.0408	0.0436	0.0586	0.0988	0.1162	0.1402	0.1356
		800	0.0224	0.0293	0.0266	0.0351	0.0502	0.0626	0.1684	0.1583
		1600	0.0173	0.0218	0.0183	0.0243	0.0351	0.0453	0.1023	0.1063
	(d)	400	0.0277	0.0431	0.0207	0.0310	0.0602	0.0800	0.0528	0.0594
		800	0.0226	0.0332	0.0061	0.0080	0.0138	0.0178	0.0581	0.0648
		1600	0.0198	0.0314	0.0018	0.0025	0.0049	0.0065	0.0154	0.0178

[1]Dark cells demonstrate what happens to bias and RMSE, if we still use (2.25) after failing to reject the null.

[2]Cells in boxes highlight the best estimates for each case in terms of RMSE.

4 Empirical Study

In this section, we provide an application of the single-index model proposed in this paper to the analysis of large commercial banks in the U.S. Compared with the conventional fully-parametric translog cost function. Model (1.1) is suitable for modelling the production technology, because the single-index setting (i.e., $g_0(x'\theta_0)$) is more flexible than the commonly-used translog linear form, which limits the variety of shapes the cost function is permitted to take.

The dataset used in this application is obtained from the Reports of Income and Condition (Call Reports) published by the Federal Reserve Bank of Chicago. More specifically, we focus on the data of 2004 only and examine large banks with assets of at least \$1 billion. To select the relevant variables, we follow the commonly-accepted intermediation approach (Sealey and Lindley, 1977). On the input side, three inputs are included: (1) the quantity of labor; (2) the quantity of purchased funds and deposits; and (3) the quantity of physical capital, which includes premises and other fixed assets. On the output side, three outputs are specified: (1) consumer loans; (2) securities, which includes all non-loan financial assets; and (3) non-consumer loans, which is composed of industrial, commercial, and real estate loans. All the quantities are constructed as in Berger and Mester (2003).

Traditionally, when one models the translog cost function in the field of production econometrics, a common approach is

$$\ln \frac{C}{w_3} = \alpha_0 + \sum_{j=1}^{2} \alpha_j \ln \frac{w_j}{w_3} + \sum_{m=1}^{3} \gamma_m \ln y_m + e, \tag{4.1}$$

where we divide C, w_1 and w_2 by w_3 to maintain linear homogeneity with respect to input prices (see Feng and Zhang (2012) for similar treatment on the homogeneity); C is total cost; y_m for $m = 1, \ldots, 3$ are variables representing outputs; and w_j for $j = 1, \ldots, 3$ are variables representing input prices. Then in order to capture the non-linearity of regressors, one can further introduce interaction terms, quadratic terms or even higher order polynomials to (4.1) (see Feng and Serletis (2008), and Feng and Zhang (2012) for some examples). Although introducing interaction terms and polynomials to the system is convenient for the purpose of modelling the non-linearity of the regressors, it is impossible to exhaust all the possible functional forms for (4.1) in practice.

In order to allow for a flexible functional form for the translog cost function, we propose using a single-index model of the form:

$$\ln \frac{C}{w_3} = C\left(\frac{w_1}{w_3},\ \frac{w_2}{w_3},\ y_1,\ y_2,\ y_3\right) + e = g_0(x'\theta_0) + e, \tag{4.2}$$

where $x = (\ln \frac{w_1}{w_3},\ \ln \frac{w_2}{w_3},\ \ln y_1,\ \ln y_2,\ \ln y_3,\ 1)'$, $C(\cdot)$ represents the normalized cost function, and e is a random error.

For the purpose of comparison and selection of an optimal truncation parameter, we first estimate (4.1) by OLS, and estimate (4.2) by (2.10) and (2.19) under different k's. For each estimation method, we calculate the root of prediction mean of squares (RPREMS) by the next formula:

$$\text{RPREMS} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (\widehat{y}_{-i} - y_i)^2}, \tag{4.3}$$

where $\widehat{y}_{-i}$ denotes the leave-one-out estimate (i.e., we implement estimation without the i^{th} individual, and then use the estimate and observation on the i^{th} individual to calculate the prediction error). More details and discussions on using RPREMS as a criterion to measure the performances of different methods can be seen in Chu et al. (2016).

Table 3: Root of Prediction Mean of Squares (RPREMS)

Model (4.1) with OLS	0.1961			
	$k = 2$	$k = 3$	$k = 4$	$k = 5$
Model (4.2) with Method (2.10)	0.1961	0.1921	0.2472	0.5462
Model (4.2) with Method (2.19)	0.1961	0.1952	0.1952	0.1958

As we can see from Table 3, the single-index setting with the method (2.10) provides the best RPREMS when $k = 2$. Generally speaking, all models and methods work well, but the single-index model fits the dataset slightly better in terms of RPREMS. This is not surprising, as (4.1) is just a special case of (4.2) by setting the truncation parameter k to 2. That is also why the three numbers in the second column of Table 3 are identical. It is noteworthy that for method (2.10) the RPREMES goes up very quickly as k increases, which matches Lemma 2.1 well and is due to the curse of dimensionality caused by expansion (2.5).

Below we report the results of the optimal estimate in terms of RPREMES (i.e., model (4.2) with method (2.10) and $k = 3$). The estimated parameters and their corresponding standard deviations are reported in the Table 4. Moreover, we plot the estimated link function $\widehat{g}(w) = 22.0732 \cdot h_0(w) - 4.2859 \cdot h_1(w) + 0.9606 \cdot h_2(w)$ (solid line) and the corresponding 95% confidence bands (dash lines) in Figure 1 based on 300 replications using the wild bootstrapping technique. Given the value of $\widehat{\theta}$ shown in Table 4, the majority of $x_i'\widehat{\theta}$ with $i = 1, \ldots, 466$ fall in the interval $[4, 5.6]$ in our data set, so we concentrate on this particular interval as we plot the estimated curve below. As shown in Figure 1, the link function (the solid line) is indeed a nonlinear curve.

Table 4: Parameter Estimates

Regressor	$\ln \frac{w_1}{2_3}$	$\ln \frac{w_2}{2_3}$	$\ln y_1$	$\ln y_2$	$\ln y_3$	constant
Est	0.6796	-0.3532	0.1792	0.2333	-0.4782	0.3134
Std	0.0412	0.0035	0.1029	0.0676	0.1102	0.0026

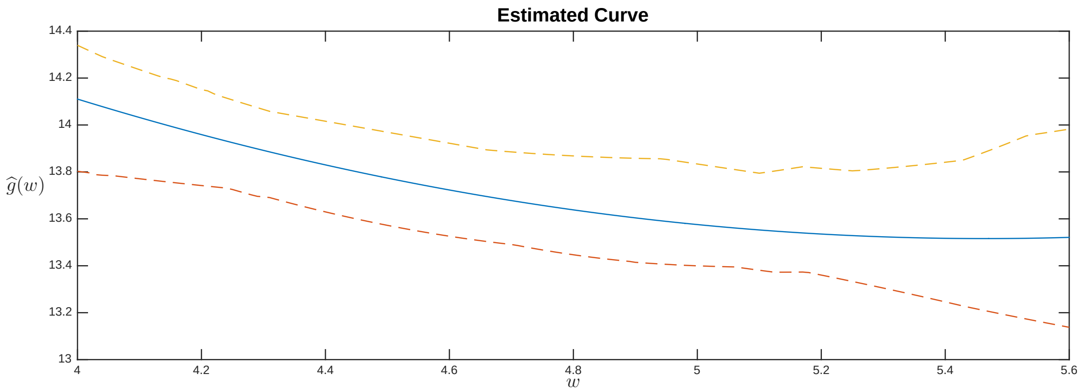

Figure 1: The Estimated Link Function $\widehat{g}(w)$

After obtaining the estimates of $\widehat{\theta}$ and $\widehat{g}(w)$, one then can follow the traditional methods used in production econometrics to calculate some interesting results like return to scales, technical change, and so forth (e.g., Feng and Zhang (2012) and Henderson et al. (2015)). We omit these results here, as they are not the main focus of this paper.

5 Conclusion

In this study, a cross-sectional single-index model $y = g_0(x'\theta_0) + e$ is considered. An overlooked question in the literature of single-index models has been solved and brought to people's attending for the first time.

Two types of estimation methods have been proposed. The first type of estimator derived from an optimization with constraint of identification condition possesses two different rates of convergence along with the direction θ_0 and that is orthogonal to θ_0. The second type of estimator is of closed-form which is derived from a simple linear model and hence relaxes the compactness condition on the parameter space. The corresponding asymptotic properties of the estimators have been established. The finite sample properties of the two estimates have been evaluated through an extensive Monte Carlo study and then an empirical application. In some special situations we have compared the asymptotic efficiency of the estimates (2.10) and (2.25), but it is still an open question in general. Additionally, we have also implemented an extensive simulation study for the purpose of comparison. To conclude, in order to get a reliable estimate in practice, one may need to find an initial estimate from (2.25), and then use the initial value to start an estimation procedure for (2.10) (or some kernel based method documented in Xia (2006)[4]).

6 Acknowledgements

The authors would like to thank seminar participants at several places for their constructive comments and suggestions on earlier versions of this paper. Thanks from the authors also go to the National Nature Science Foundation of China under project 71671143 and the Australian Research Council Discovery Grants Program for its support under Grant numbers: DP150101012 & DP170104421.

Appendix A: Proofs

Lemma A.1. *Let $g^{(i)}(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ where $0 \le i \le r$ and $r \ge 1$. Then $\|\delta_k(w)\|_{L^2}^2 = O(k^{-r})$ and $\|\delta_k'(w)\|_{L^2}^2 = o(1)$ as $k \to \infty$.*

Proof of Lemma A.1:

By definition, $H_j(w) = (-1)^j e^{w^2/2}(e^{-w^2/2})^{(j)}$, and integration by parts we have

$$\begin{aligned} c_j(g) =& \int g(w)h_j(w)e^{-w^2/2}dw = \frac{(-1)^j}{b_j}\int g(w)d(e^{-w^2/2})^{(j-1)} \\ =& \frac{(-1)^j}{b_j}\int g^{(1)}(w)(e^{-w^2/2})^{(j-1)}dw = -\frac{b_{j-1}}{b_j}\frac{(-1)^{j-1}}{b_{j-1}}\int g^{(1)}(w)h_{j-1}(w)e^{-w^2/2}dw \\ =& -\frac{1}{\sqrt{j}}c_{j-1}(g^{(1)}), \end{aligned}$$

where $b_j = \sqrt{j!}$ and $c_{j-1}(g^{(1)})$ is the coefficient of the expansion of $g^{(1)}(w)$. Hence, it follows from induction that for $j \ge r$,

$$c_j(g) = (-1)^r \frac{1}{\sqrt{j \cdots (j-r)}} c_{j-r}(g^{(r)}). \tag{A.1}$$

Observe that, for $k > r$,

$$\begin{aligned} \|\gamma_k(w)\|_{L^2} =& \int \gamma_k^2(w)e^{-w^2/2}dw = \sum_{j=k}^{\infty} c_j^2(g) = \sum_{j=k}^{\infty}\frac{1}{j\cdots(j-r)}c_{j-r}^2(g^{(r)}) \\ \le& O(k^{-r})\sum_{j=k}^{\infty} c_{j-r}^2(g^{(r)}) = o(k^{-r}), \end{aligned}$$

since $\sum_{j=k}^{\infty} c_{j-r}^2(g^{(r)}) = o(1)$ due to Parseval's equality $\sum_{j=r}^{\infty} c_{j-r}^2(g^{(r)}) = \|g^{(r)}(w)\|_{L^2}^2 < \infty$.

In addition, noting that $h_j'(w) = \sqrt{j}h_{j-1}(w)$,

$$\|\gamma_k^{(1)}(w)\|_{L^2} = \int |\gamma_k^{(1)}(w)|^2 e^{-w^2/2}dw = \sum_{j=k}^{\infty} c_j^2(g)j = \sum_{j=k}^{\infty}\frac{1}{j}c_{j-1}^2(g^{(1)})j = \sum_{j=k}^{\infty} c_{j-1}^2(g^{(1)}) = o(1).$$

The proof is then completed. ■

[4]It is noteworthy that the two algorithms studied in Xia (2006) both have requirements on the initial estimates.

Proof of Theorem 2.1:

It is easy to see $W_{N,\lambda}(\theta, g)$ and $\mathcal{W}_{N,\lambda}(\theta, C_k)$ defined in (2.8) and (2.9) are interchangeable (i.e., $L_N(\theta, g)$ and $\mathcal{L}_N(\theta, C_k)$ are interchangeable). Furthermore, note the term $\lambda(\|\theta\|^2 - 1)$ in (2.8) and (2.9) are independent of the sample $\{y_i, x_i\}_{i=1}^N$. Thus, we focus on $L_N(\theta, g)$ first and show $\max_{\Theta\times G}|L_N(\theta, g) - L(\theta, g)| \to_P 0$ by using Lemma A2 of Newey and Powell (2003).

Condition (1) of Lemma A2 of Newey and Powell (2003) holds according to Assumption 1.2. By Weak Law of Large Numbers (WLLN), it is easy to know $L_N(\theta, g) = L(\theta, g)(1 + o_P(1))$, which indicates that condition (ii) of Lemma A2 of Newey and Powell (2003) holds. We then just need to show the continuity of $L_1(\theta, g)$, which then indicates that the continuity of $L_N(\theta, g)$ holds with probability approaching to 1 (i.e., condition (iii) of Lemma A2 of Newey and Powell (2003) holds). Further note that $L(\theta, g) = L_1(\theta, g) + \sigma_e^2$, where $L_1(\theta, g) = E[g_0(x_1'\theta_0) - g(x_1'\theta)]^2$. Thus, we just need to focus on the continuity of $L_1(\theta, g)$ below. To do so, for any given (θ_1, g_1) and (θ_2, g_2) belonging to $\Theta \times G$, write

$$|L_1(\theta_1, g_1) - L_1(\theta_2, g_2)| \le |L_1(\theta_1, g_1) - L_1(\theta_1, g_2)| + |L_1(\theta_1, g_2) - L_1(\theta_2, g_2)|. \tag{A.2}$$

For the first term on right hand side (RHS) of (A.2), write

$$\begin{aligned}
&|L_1(\theta_1, g_1) - L_1(\theta_1, g_2)| \\
=& \left|E\left[(g_0(x_1'\theta_0) - g_1(x_1'\theta_1))^2 - (g_0(x_1'\theta_0) - g_2(x_1'\theta_1))^2\right]\right| \\
=& \left|E\left[(g_2(x_1'\theta_1) - g_1(x_1'\theta_1)) \cdot (2g_0(x_1'\theta_0) - g_1(x_1'\theta_1) - g_2(x_1'\theta_1))\right]\right| \\
\le& \left\{E\left[g_2(x_1'\theta_1) - g_1(x_1'\theta_1)\right]^2 \cdot E\left[2g_0(x_1'\theta_0) - g_1(x_1'\theta_1) - g_2(x_1'\theta_1)\right]^2\right\}^{1/2},
\end{aligned}$$

where the inequality follows from Cauchy-Schwarz inequality. We then focus on

$$E\left[g_2(x_1'\theta_1) - g_1(x_1'\theta_1)\right]^2 \quad \text{and} \quad E\left[2g_0(x_1'\theta_0) - g_1(x_1'\theta_1) - g_2(x_1'\theta_1)\right]^2$$

respectively. Note that

$$\begin{aligned}
E\left[g_2(x_1'\theta_1) - g_1(x_1'\theta_1)\right]^2 &= \int (g_1(w) - g_2(w))^2 f_{\theta_1}(w)dw \\
= \int (g_1(w) - g_2(w))^2 &\exp(-w^2/2) \cdot \exp(w^2/2) f_{\theta_1}(w)dw \\
\le O(1) \int (g_1(w) - g_2(w))^2 &\exp(-w^2/2)dw = O(1)\|g_1 - g_2\|_{L^2}^2,
\end{aligned}$$

where the inequality follows from Assumption 1.2.

For $\forall(\theta, g) \in \Theta \times G$,

$$\begin{aligned}
E[g(x_1'\theta)]^2 &= \int (g(w))^2 f_\theta(w)dw = \int (g(w))^2 \exp(-w^2/2) \cdot \exp(w^2/2) f_\theta(w)dw \\
&\le O(1) \int (g(w))^2 \exp(-w^2/2)dw = O(1)\|g\|_{L^2}^2 \le O(1),
\end{aligned}$$

where the first inequality follows from Assumption 1.2, and the second inequality follows from $\Theta \times G$ being compact set. Similarly, we have

$$\begin{aligned}
&E\left[2g_0(x_1'\theta_0) - g_2(x_1'\theta_1) - g_1(x_1'\theta_2)\right]^2 \\
\le& 8E\left[g_0(x_1'\theta_0)\right]^2 + 4E\left[g_2(x_1'\theta_1)\right]^2 + 4E\left[g_1(x_1'\theta_2)\right]^2 \le O(1).
\end{aligned}$$

Then, we have shown

$$|L_1(\theta_1, g_1) - L_1(\theta_1, g_2)| \le O(1)\|g_1 - g_2\|_{L^2}. \tag{A.3}$$

We now consider the second term on RHS of (A.2):

$$
\begin{aligned}
&|L_1(\theta_1, g_2) - L_1(\theta_2, g_2)| \\
&= \left| E\left[(g_0(x_1'\theta_0) - g_2(x_1'\theta_1))^2 - (g_0(x_1'\theta_0) - g_2(x_1'\theta_2))^2 \right] \right| \\
&= |E\left[(g_2(x_1'\theta_2) - g_2(x_1'\theta_1)) \cdot (2g_0(x_1'\theta_0) - g_2(x_1'\theta_1) - g_2(x_1'\theta_2)) \right]| \\
&\leq \left\{ E\left[g_2(x_1'\theta_2) - g_2(x_1'\theta_1) \right]^2 \cdot E\left[2g_0(x_1'\theta_0) - g_2(x_1'\theta_1) - g_2(x_1'\theta_2) \right]^2 \right\}^{1/2}.
\end{aligned}
$$

Similar to the above, it is easy to show $E\left[2g_0(x_1'\theta_0) - g_2(x_1'\theta_1) - g_2(x_1'\theta_2)\right]^2$ is bounded uniformly on $\Theta \times G$, so we just need to focus on $E\left[g_2(x_1'\theta_2) - g_2(x_1'\theta_1)\right]^2$. Write

$$
\begin{aligned}
&E\left[g_2(x_1'\theta_2) - g_2(x_1'\theta_1) \right]^2 = E\left[(\theta_2 - \theta_1)' x_1 x_1' (\theta_2 - \theta_1) \{ g_2^{(1)}(x_1'\theta^*) \}^2 \right] \\
&\leq \|\theta_2 - \theta_1\|^2 E\left\| x_1 x_1' \{ g_2^{(1)}(x_1'\theta^*) \}^2 \right\| \leq O(1) \|\theta_2 - \theta_1\|^2,
\end{aligned}
$$

where θ^* lies between θ_1 and θ_2, and the second inequality follows from Assumption 1.2. Therefore, we have

$$
|L_1(\theta_1, g_2) - L_1(\theta_2, g_2)| \leq O(1) \|\theta_2 - \theta_1\|. \tag{A.4}
$$

By (A.3), (A.4) and the fact that $L(\theta, g) = L_1(\theta, g) + \sigma_e^2$, we immediately obtain

$$
|L(\theta_1, g_1) - L(\theta_2, g_2)| \leq O(1) \left\| (\theta_1, g_1) - (\theta_2, g_2) \right\|_2,
$$

which indicates the continuity of $L(\theta, g)$. Then we have shown $\max_{\Theta \times G} |L_N(\theta, g) - L(\theta, g)| \to_P 0$.

It is easy to see that for (2.8), $W_{N,\lambda}(\theta_0, g_0) = \sigma_e^2 + o_P(1)$ regardless of the value of λ. If $(\widetilde{\theta}, \widetilde{g}) \not\to_P (\theta_0, g_0)$, then we have $W_{N,\lambda}(\widetilde{\theta}, \widetilde{g}) > \sigma_e^2$ with probability approaching 1 based on the above analysis, which violates the definition of (2.10). Therefore, we must have $(\widetilde{\theta}, \widetilde{g}) \to_P (\theta_0, g_0)$. The proof is now completed. ■

Remark 6. *By definition of (2.10), we have*

$$
0 = \frac{\partial}{\partial \theta} \mathcal{W}_{N,\widetilde{\lambda}}(\widetilde{\theta}, \widetilde{C}_k), \quad 0 = \frac{\partial}{\partial C_k} \mathcal{W}_{N,\widetilde{\lambda}}(\widetilde{\theta}, \widetilde{C}_k), \quad 0 = \frac{\partial}{\partial \lambda} \mathcal{W}_{N,\widetilde{\lambda}}(\widetilde{\theta}, \widetilde{C}_k).
$$

Note that $0 = \frac{\partial}{\partial \lambda} \mathcal{W}_{N,\widetilde{\lambda}}(\widetilde{\theta}, \widetilde{C}_k)$ *gives* $\|\widetilde{\theta}\|^2 - 1 = 0$. *Thus, multiplying* $\widetilde{\theta}'$ *for both sides of* $0 = \frac{\partial}{\partial \theta} \mathcal{W}_{N,\widetilde{\lambda}}(\widetilde{\theta}, \widetilde{C}_k)$ *immediately gives*

$$
\widetilde{\lambda} = \frac{1}{N} \sum_{i=1}^{N} \left[y_i - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \widetilde{g}_k^{(1)}(x_i'\widetilde{\theta}) x_i'\widetilde{\theta}. \tag{A.5}
$$

According to Assumptions 1 and 2, we provide the next lemma before prove the asymptotic normality for $\widetilde{\theta}$.

Lemma A.2. *Suppose Assumptions 1 and 2.1-2.3 hold. Let* $V = E\left[\left(g_0^{(1)}(x_1'\theta_0) \right)^2 x_1 x_1' \right]$. *For* $\mathcal{W}_{N,\lambda}(\theta, C_k)$ *denoted in (2.9), as* $N \to \infty$,

1. $\frac{\partial^2 \mathcal{W}_{N,\lambda}}{\partial \theta \partial \theta'} \Big|_{(\theta, C_k, \lambda) = (\overline{\theta}, \widetilde{C}_k, \widetilde{\lambda})} \to_P 2V$, *where* $\overline{\theta}$ *lies between* θ_0 *and* $\widetilde{\theta}$;
2. *In addition, let Assumption 2.4 hold.*

$$
\sqrt{N} P_1' \frac{\partial \mathcal{W}_{N,\lambda}}{\partial \theta} \Big|_{(\theta, C_k, \lambda) = (\theta_0, \widetilde{C}_k, \widetilde{\lambda})} + P_1' \sqrt{N} 2 S_N(\widetilde{g}_k) \to_D N(0, 4\sigma_e^2 P_1' V P_1),
$$

where $S_N(\widetilde{g}_k) = \frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\theta_0) \right] \widetilde{g}_k^{(1)}(x_i'\theta_0) x_i$ *and* P_1 *is defined above (2.11).*

Proof of Lemma A.2:

Since $\overline{\theta}$ lies between θ_0 and $\widetilde{\theta}$, it is easy to know $\|(\overline{\theta},\widetilde{g}_k)-(\theta_0,g_0)\|_2\to_P 0$ by Theorem 2.1. Thus, it is reasonable to focus on a sufficiently small neighbourhood of (θ_0,g_0) in the following proof. Then the analysis is similar to the arguments of (2.19)-(2.21) of Amemiya (1993), and part (b) of Lemma B.1 of Yu and Ruppert (2002). Note that one can easily extend the arguments (2.19)-(2.21) of Amemiya (1993) to the current setting by treating β and $\|\cdot\|$ of Amemiya (1993) as (θ,g) and $\|\cdot\|_2$ of this paper respectively, which then becomes exactly the same as Lemma A.2 of Newey and Powell (2003). It then allows us to write, for example,

$$\frac{1}{N}\sum_{i=1}^N\left(g^{(1)}(x_i'\theta)\right)^2x_ix_i'\Big|_{(\theta,g)=(\overline{\theta},\widetilde{g}_k)}=E\left[\left(g^{(1)}(x_1'\theta)\right)^2x_1x_1'\right]\Big|_{(\theta,g)=(\overline{\theta},\widetilde{g}_k)}+o_P(1)$$

by Assumption 2.2, which further allows us to simplify the analysis by focusing on the expectation like the right hand side above (c.f. Lemma A.2 of Newey and Powell (2003)). Also, in the following derivations, we will repeatedly use notation $g_k(w)=\mathcal{H}(w)'C_k$, which has been defined in (2.3).

(1). We now start focusing on $\frac{\partial^2\mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'}|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})}$.

$$\begin{aligned}
&\frac{\partial^2\mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'}\Big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})}\\
&=\left\{\frac{2}{N}\sum_{i=1}^N\left(g_k^{(1)}(x_i'\theta)'\right)^2x_ix_i'-\frac{2}{N}\sum_{i=1}^N\left(y_i-g_k(x_i'\theta)\right)g_k^{(2)}(x_i'\theta)x_ix_i'+2\lambda I_d\right\}\Big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})}\\
&:=2J_{1,N}\Big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}-2J_{2,N}\Big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}+2\lambda I_d\Big|_{\lambda=\widetilde{\lambda}},
\end{aligned}\tag{A.6}$$

where the definitions of $J_{1,N}$ and $J_{2,N}$ should be obvious, and g_k is denoted in (2.3).

For $J_{1,N}$, write

$$\begin{aligned}
J_{1,N}&=\frac{1}{N}\sum_{i=1}^N\left(g_k^{(1)}(x_i'\theta)\right)^2x_ix_i'=\frac{1}{N}\sum_{i=1}^N\left(g_0^{(1)}(x_i'\theta_0)\right)^2x_ix_i' &\text{(A.7)}\\
&+\frac{1}{N}\sum_{i=1}^N\left(g_k^{(1)}(x_i'\theta)'\right)^2x_ix_i'-\frac{1}{N}\sum_{i=1}^N\left(g_{0,k}^{(1)}(x_i'\theta)\right)^2x_ix_i'\\
&+\frac{1}{N}\sum_{i=1}^N\left(g_{0,k}^{(1)}(x_i'\theta)\right)^2x_ix_i'-\frac{1}{N}\sum_{i=1}^N\left(g_0^{(1)}(x_i'\theta)\right)^2x_ix_i'\\
&+\frac{1}{N}\sum_{i=1}^N\left(g_0^{(1)}(x_i'\theta)\right)^2x_ix_i'-\frac{1}{N}\sum_{i=1}^N\left(g_0^{(1)}(x_i'\theta_0)\right)^2x_ix_i\\
&:=\frac{1}{N}\sum_{i=1}^N\left(g_0^{(1)}(x_i'\theta_0)\right)^2x_ix_i'+J_{11,N}+J_{12,N}+J_{13,N}, &\text{(A.8)}
\end{aligned}$$

where $g_{0,k}$ is denoted in (2.4); the definitions of $J_{11,N}$, $J_{12,N}$ and $J_{13,N}$ should be obvious. By Assumption 2.2, we know

$$\frac{1}{N}\sum_{i=1}^N\left(g_0^{(1)}(x_i'\theta_0)\right)^2x_ix_i'=E\left[\left(g_0^{(1)}(x_1'\theta_0)\right)^2x_1x_1'\right]+o_P(1).$$

We then focus on $J_{11,N}$ to $J_{13,N}$ respectively. For $J_{11,N}$, we write

$$\begin{aligned}
&\|J_{11,N}\|\,|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}\\
&=\left\|\frac{1}{N}\sum_{i=1}^N\left(\dot{\mathcal{H}}(x_i'\theta)'C\right)^2x_ix_i'-\frac{1}{N}\sum_{i=1}^N\left(\dot{\mathcal{H}}(x_i'\theta)'C_{0,2:k}\right)^2x_ix_i'\right\|\,\Big|_{(\theta,C)=(\overline{\theta},\widetilde{C}_{2:k})}\\
&\le\left\|E\left[\left(\dot{\mathcal{H}}(x_1'\theta)'C\right)^2x_1x_1'\right]-E\left[\left(\dot{\mathcal{H}}(x_1'\theta)'C_{0,2:k}\right)^2x_1x_1'\right]\right\|\,\Big|_{(\theta,C)=(\overline{\theta},\widetilde{C}_{2:k})}+o_P(1)\\
&=\left\|E\left[(C-C_{0,2:k})'\dot{\mathcal{H}}(x_1'\theta)\dot{\mathcal{H}}(x_1'\theta)'(C+C_{0,2:k})x_1x_1'\right]\right\|\,\Big|_{(\theta,C)=(\overline{\theta},\widetilde{C}_{2:k})}+o_P(1)
\end{aligned}$$

$$
\begin{aligned}
&\leq \left\{ E \left| (C - C_{0,2:k})' \dot{\mathcal{H}}(x_1'\theta) \right|^2 E \left\| \dot{\mathcal{H}}(x_1'\theta)'(C + C_{0,2:k}) x_1 x_1' \right\|^2 \right\}^{1/2} \Big|_{(\theta, C) = (\overline{\theta}, \widetilde{C}_{2:k})} + o_P(1) \\
&\leq \left\{ (C - C_{0,2:k})' E \left[\dot{\mathcal{H}}(x_1'\theta) \dot{\mathcal{H}}(x_1'\theta)' \right] (C - C_{0,2:k}) \right\}^{1/2} \\
&\quad \cdot \left\{ 2E \left\| g_k^{(1)}(x_1'\theta) x x \right\|^2 + 2E \left\| g_{0,k}^{(1)}(x_1'\theta) x_1 x_1' \right\|^2 \right\}^{1/2} \Big|_{(\theta, C) = (\overline{\theta}, \widetilde{C}_{2:k})} + o_P(1) \\
&\leq O(1) \| \widetilde{C}_{2:k} - C_{0,2:k} \| + o_P(1) \leq O(1) \| \widetilde{g}_k - g_0 \|_{L^2} + o_P(1) = o_P(1),
\end{aligned}
$$

where $C_{0,2:k}$ and $\widetilde{C}_{2:k}$ denote the vectors consisting of the 2^{nd} to k^{th} elements of $C_{0,k}$ and $\widetilde{C}_k$respectively, the first inequality follows from Assumption 2.2, the second inequality follows from Cauchy-Schwarz inequality, the fourth inequality follows from Assumptions 1.2 and 2.3, and the last equality follows from Theorem 2.1.

For $J_{12,N}$, write

$$
\begin{aligned}
\| J_{12,N} \| \, |_{\theta = \overline{\theta}} &= \left\| \frac{1}{N} \sum_{i=1}^N \left(g_{0,k}^{(1)}(x_i'\theta) \right)^2 x_i x_i' - \frac{1}{N} \sum_{i=1}^N \left(g_0^{(1)}(x_i'\theta) \right)^2 x_i x_i' \right\| \Big|_{\theta = \overline{\theta}} \\
&\leq \left\| E \left[\left\{ \left(g_{0,k}^{(1)}(x_1'\theta) \right)^2 - \left(g_0^{(1)}(x_1'\theta) \right)^2 \right\} x_1 x_1' \right] \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&= \left\| E \left[\left(g_{0,k}^{(1)}(x_1'\theta) - g_0^{(1)}(x_1'\theta) \right) \left(g_{0,k}^{(1)}(x_1'\theta) + g_0^{(1)}(x_1'\theta) \right) x_1 x_1' \right] \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&= \left\| E \left[\delta_{0,k}^{(1)}(x_1'\theta) \left(g_{0,k}^{(1)}(x_1'\theta) + g_0^{(1)}(x_1'\theta) \right) x_1 x_1' \right] \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&\leq E \left\| \delta_{0,k}^{(1)}(x_1'\theta) g_{0,k}^{(1)}(x_1'\theta) x_1 x_1' \right\| \Big|_{\theta = \overline{\theta}} + E \left\| \delta_{0,k}^{(1)}(x_1'\theta) g_0^{(1)}(x_1'\theta) x_1 x_1' \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&\leq \left\{ E \left| \delta_{0,k}^{(1)}(x_1'\theta) \right|^2 E \left\| g_{0,k}^{(1)}(x_1'\theta) x_1 x_1' \right\|^2 \right\}^{1/2} \Big|_{\theta = \overline{\theta}} \\
&\quad + \left\{ E \left| \delta_{0,k}^{(1)}(x_1'\theta) \right|^2 E \left\| g_0^{(1)}(x_1'\theta) x_1 x_1' \right\|^2 \right\}^{1/2} \Big|_{\theta = \overline{\theta}} + o_P(1) = o_P(1),
\end{aligned}
$$

where the first inequality follows from Assumption 2.2, the third equality follows from (2.4), the third inequality follows from Cauchy-Schwarz inequality, and the last equality follows from Assumption 1.2 and Lemma A.1.

For $J_{13,N}$, write

$$
\begin{aligned}
\| J_{13,N} \| \, |_{\theta = \overline{\theta}} &= \left\| \frac{1}{N} \sum_{i=1}^N \left(g_0^{(1)}(x_i'\theta) \right)^2 x_i x_i' - \frac{1}{N} \sum_{i=1}^N \left(g_0^{(1)}(x_i'\theta_0) \right)^2 x_i x_i' \right\| \Big|_{\theta = \overline{\theta}} \\
&\leq \left\| E \left[\left(g_0^{(1)}(x_1'\theta) \right)^2 x_1 x_1' \right] - E \left[\left(g_0^{(1)}(x_1'\theta_0) \right)^2 x_1 x_1' \right] \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&= \left\| E \left[\left(g_0^{(1)}(x_1'\theta) - g_0^{(1)}(x_1'\theta_0) \right) \left(g_0^{(1)}(x_1'\theta) + g_0^{(1)}(x_1'\theta_0) \right) x_1 x_1' \right] \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&= \left\| E \left[g_0^{(2)}(x_1'\theta^*) \left(x_1'\theta - x_1'\theta_0 \right) \left(g_0^{(1)}(x_1'\theta) + g_0^{(1)}(x_1'\theta_0) \right) x_1 x_1' \right] \right\| \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&\leq \left\{ \| \theta - \theta_0 \| E \left[\left\| g_0^{(2)}(x_1'\theta^*) x_1 \right\| \left\| \left(g_0^{(1)}(x_1'\theta) + g_0^{(1)}(x_1'\theta_0) \right) x_1 x_1' \right\| \right] \right\} \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&\leq \left[\| \theta - \theta_0 \| \left\{ E \left\| g_0^{(2)}(x_1'\theta^*) x_1 \right\|^2 E \left\| g_0^{(1)}(x_1'\theta) x_1 x_1' \right\|^2 \right\}^{1/2} \right] \Big|_{\theta = \overline{\theta}} + o_P(1) \\
&\quad + \left[\| \theta - \theta_0 \| \left\{ E \left\| g_0^{(2)}(x_1'\theta^*) x_1 \right\|^2 E \left\| g_0^{(1)}(x_1'\theta_0) x_1 x_1' \right\|^2 \right\}^{1/2} \right] \Big|_{\theta = \overline{\theta}} + o_P(1) = o_P(1),
\end{aligned}
$$

where θ^* lies between θ and θ_0, the first inequality follows from Assumption 2.2, the second equality follows from Mean Value Theorem, the third inequality follows from triangular inequality and Cauchy-Schwarz inequality, and the last equality follows from Assumption 1.2 and the fact that $\|\overline{\theta} - \theta_0\| \to_P 0$.

Based on the above derivations, we have shown $J_{1,N} \mid_{(\theta, C) = (\overline{\theta}, \widetilde{C}_k)} \to_P V$. We now turn to $J_{2,N}$ and write

$$
J_{2,N} \mid_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)}
$$

$$= \left\{ \frac{1}{N} \sum_{i=1}^{N} e_i g_k^{(2)}(x_i'\theta) x_i x_i' + \frac{1}{N} \sum_{i=1}^{N} \left(g_0(x_i'\theta_0) - g_k(x_i'\theta) \right) g_k^{(2)}(x_i'\theta) x_i x_i' \right\} \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)}$$
$$:= J_{21,N}\Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + J_{22,N}\Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)},$$

where the definitions of $J_{21,N}$ and $J_{22,N}$ should be obvious.

By Assumption 1 and the proof similar to that given in Theorem 2.1, it is easy to show that $J_{21,N} = O_P\left(\frac{1}{\sqrt{N}}\right)$ uniformly.

For $J_{22,N}$, write

$$\begin{aligned} J_{22,N} &= \frac{1}{N} \sum_{i=1}^{N} g_0(x_i'\theta_0) g_k^{(2)}(x_i'\theta) x_i x_i' - \frac{1}{N} \sum_{i=1}^{N} g_0(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' \\ &\quad + \frac{1}{N} \sum_{i=1}^{N} g_0(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' - \frac{1}{N} \sum_{i=1}^{N} g_{0,k}(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' \\ &\quad + \frac{1}{N} \sum_{i=1}^{N} g_{0,k}(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' - \frac{1}{N} \sum_{i=1}^{N} g_k(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' \\ &:= J_{221,N} + J_{222,N} + J_{223,N}. \end{aligned}$$

For $J_{221,N}$, write

$$\begin{aligned} &\|J_{221,N}\| |_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} \\ &= \left\| \frac{1}{N} \sum_{i=1}^{N} g_0(x_i'\theta_0) g_k^{(2)}(x_i'\theta) x_i x_i' - \frac{1}{N} \sum_{i=1}^{N} g_0(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' \right\| \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} \\ &\leq \left\| E\left[\left(g_0(x_1'\theta_0) - g_0(x_1'\theta) \right) g_k^{(2)}(x_1'\theta) x_1 x_1' \right] \right\| \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + o_P(1) \\ &= \left\| E\left[g_0^{(1)}(x_1'\theta^*)(x_1'\theta_0 - x'\theta) g_k^{(2)}(x_1'\theta) x_1 x_1' \right] \right\| \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + o_P(1) \\ &\leq \left[\|\theta_0 - \overline{\theta}\| \left\{ E\left\| g_0^{(1)}(x_1'\theta^*) x_1 x_1' \right\|^2 E\left\| g_k^{(2)}(x_1'\theta) x_1 \right\|^2 \right\}^{1/2} \right] \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + o_P(1) \\ &= o_P(1), \end{aligned}$$

where θ^* lies between θ and θ_0, the first inequality follows from Assumption 2.2, the second equality follows from Mean Value Theorem, the third inequality follows from Cauchy-Schwarz inequality, and the last equality follows from Assumption 1.2 and the fact that $\|\overline{\theta} - \theta_0\| \to_P 0$.

For $J_{222,N}$, write

$$\begin{aligned} &\|J_{222,N}\| |_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} \\ &= \left\| \frac{1}{N} \sum_{i=1}^{N} g_0(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' - \frac{1}{N} \sum_{i=1}^{N} g_{0,k}(x_i'\theta) g_k^{(2)}(x_i'\theta) x_i x_i' \right\| \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} \\ &\leq \left\| E\left[g_0(x_1'\theta) g_k^{(2)}(x_1'\theta) x_1 x_1' - g_{0,k}(x_1'\theta) g_k^{(2)}(x_1'\theta) x_1 x_1' \right] \right\| \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + o_P(1) \\ &= \left\| E\left[\delta_{0,k}(x_1'\theta) g_k^{(2)}(x_1'\theta) x_1 x_1' \right] \right\| \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + o_P(1) \\ &\leq \left\{ E\left| \delta_{0,k}(x_1'\theta) \right|^2 E\left\| g_k^{(2)}(x_1'\theta) x_1 x_1' \right\|^2 \right\}^{1/2} \Big|_{(\theta, C_k) = (\overline{\theta}, \widetilde{C}_k)} + o_P(1) = o_P(1), \end{aligned}$$

where the first inequality follows from Assumption 2.2, the second inequality follows from Cauchy-Schwarz inequality, and the last equality follows from Assumption 2.2 and Lemma A.1.

For $J_{223,N}$, write

$$\begin{aligned}
&\|J_{223,N}\|\big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}\\
&=\left\|\frac{1}{N}\sum_{i=1}^{N}g_{0,k}(x_i'\theta)g_k^{(2)}(x_i'\theta)x_ix_i'-\frac{1}{N}\sum_{i=1}^{N}g_k(x_i'\theta)g_k^{(2)}(x_i'\theta)x_ix_i'\right\|\Big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}\\
&\leq\left\|E\left[\mathcal{H}(x_1'\theta)'C_{0,k}g_k^{(2)}(x_1'\theta)x_1x_1'-\mathcal{H}(x'\theta)'C_kg_k^{(2)}(x'\theta)x_1x_1'\right]\right\|\Big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}+o_P(1)\\
&=\left\|E\left[\left(\mathcal{H}(x_1'\theta)'C_{0,k}-\mathcal{H}(x_1'\theta)'C_k\right)g_k^{(2)}(x_1'\theta)x_1x_1'\right]\right\|\Big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}+o_P(1)\\
&\leq\left\{(C_{0,k}-C_k)'E\left[\mathcal{H}(x_1'\theta)\mathcal{H}(x_1'\theta)'\right](C_{0,k}-C_k)E\left\|g_k^{(2)}(x_1'\theta)x_1x_1'\right\|^2\right\}^{1/2}\Big|_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}+o_P(1)\\
&\leq O(1)\|C_{0,k}-\widetilde{C}\|+o_P(1)\leq O(1)\|\widetilde{g}_k-g_0\|_{L^2}+o_P(1)=o_P(1),
\end{aligned}$$

where the first inequality follows from Assumption 2.2, the second inequality follows from Cauchy-Schwarz inequality, the third inequality follows from (5) of Lemma A.4, and the fourth inequality follows from the definition of $\|\cdot\|_{L^2}$.

Based on the derivations for $J_{221,N}$ to $J_{223,N}$, we obtain $J_{22,N}\ |_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}=o_P(1)$. In connection with $J_{21,N}\ |_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}=o_P(1)$, we have $J_{2,N}\ |_{(\theta,C_k)=(\overline{\theta},\widetilde{C}_k)}=o_P(1)$.

Similarly, we can show $\widetilde{\lambda}=o_P(1)$. Therefore, we have concluded $\frac{\partial^2\mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'}\big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})}\to_P 2V$, which immediately implies the first result of this lemma.

(2). Expand $\sqrt{N}\frac{\partial\mathcal{W}_{N,\lambda}}{\partial\theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})}$ as follows:

$$\begin{aligned}
&\sqrt{N}\frac{\partial\mathcal{W}_{N,\lambda}}{\partial\theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})}=-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[y_i-\widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i+2\sqrt{N}\widetilde{\lambda}\theta_0\\
&=-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_i\,\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0)-\widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i+2\sqrt{N}\widetilde{\lambda}\theta_0\\
&=-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_ig_0^{(1)}(x_i'\theta_0)x_i-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_i\left(\widetilde{g}_k^{(1)}(x_i'\theta_0)-g_{0,k}^{(1)}(x_i'\theta_0)-\delta_{0,k}^{(1)}(x_i'\theta_0)\right)x_i\\
&\quad-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0)-\widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i+\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[y_i-\widetilde{g}_k(x_i'\widetilde{\theta})\right]\widetilde{g}_k^{(1)}(x_i'\widetilde{\theta})x_i'\widetilde{\theta}\theta_0\\
&=-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_ig_0^{(1)}(x_i'\theta_0)x_i-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_i\left(\widetilde{g}_k^{(1)}(x_i'\theta_0)-g_{0,k}^{(1)}(x_i'\theta_0)-\delta_{0,k}^{(1)}(x_i'\theta_0)\right)x_i\\
&\quad-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0)-\widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i+\theta_0\widetilde{\theta}'\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[y_i-\widetilde{g}_k(x_i'\widetilde{\theta})\right]\widetilde{g}_k^{(1)}(x_i'\widetilde{\theta})x_i\\
&=(-I+\theta_0\widetilde{\theta}')\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_ig_0^{(1)}(x_i'\theta_0)x_i\\
&\quad-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_i\left(\widetilde{g}_k^{(1)}(x_i'\theta_0)-g_{0,k}^{(1)}(x_i'\theta_0)\right)x_i+\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_i\delta_{0,k}^{(1)}(x_i'\theta_0)x_i\\
&\quad-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[g_{0,k}(x_i'\theta_0)-\widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i-\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\delta_{0,k}(x_i'\theta_0)\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i\\
&\quad+\theta_0\widetilde{\theta}'\frac{2}{\sqrt{N}}\sum_{i=1}^{N}e_i\left[\widetilde{g}_k^{(1)}(x_i'\widetilde{\theta})-g_0^{(1)}(x_i'\theta_0)\right]x_i\\
&\quad+\theta_0\widetilde{\theta}'\frac{2}{\sqrt{N}}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0)-\widetilde{g}_k(x_i'\widetilde{\theta})\right]\widetilde{g}_k^{(1)}(x_i'\widetilde{\theta})x_i\\
&=2(-I_d+\theta_0\widetilde{\theta}')J_{1N}-2J_{2N}+2J_{3N}-2J_{4N}-2J_{5N}+2\theta_0\widetilde{\theta}'J_{6N}+2\theta_0\widetilde{\theta}'J_{7N},
\end{aligned}\tag{A.9}$$

where the definitions of J_{1N}-J_{6N} should be obvious.

Below we shall consider the terms on RHS of (A.9) one by one. By Lindeberg-Lévy CLT, it is easy to show that $J_{1N} \to_D N(0, \sigma_e^2 V)$.

For J_{2N}, consider the next expression

$$\begin{aligned}
&E\left\| \frac{2}{\sqrt{N}} \sum_{i=1}^{N} e_i \left(g_k^{(1)}(x_i'\theta_0) - g_{0,k}^{(1)}(x_i'\theta_0) \right) x_i \right\|^2 \\
&= \frac{4}{N} \sum_{i=1}^{N} \sigma_e^2 E\left\| \dot{\mathcal{H}}(x_i'\theta_0)' \left(C_{2:k} - C_{0,2:k}\right) x_i \right\|^2 \\
&= \frac{4}{N} \sum_{i=1}^{N} \sigma_e^2 \left(C_{2:k} - C_{0,2:k}\right)' E\left[\dot{\mathcal{H}}(x_i'\theta_0) \dot{\mathcal{H}}(x_i'\theta_0)' \|x_i\|^2 \right] \left(C_{2:k} - C_{0,2:k}\right) \\
&\leq O(1)\|C_k - C_{0,k}\|^2 \leq O(1)\|g_k - g_0\|_{L^2}^2,
\end{aligned}$$

where $C_{2:k}$ and $C_{0,2:k}$ define the vectors consisting of the 2^{nd} to k^{th} elements of C_k and $C_{0,k}$, respectively, and the first inequality follows from Assumption 2.3. In connection with Theorem 2.1 and the arguments similar to those given in Remark 3, we obtain $J_{2N} = o_P(1)$. Similarly, $J_{6N} = o_P(1)$.

For J_{3N}, we have

$$\begin{aligned}
&E\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^{N} e_i \delta_{0,k}^{(1)}(x_i'\theta_0) x_i \right\|^2 \leq O(1) E\left\| e_1 \delta_{0,k}^{(1)}(x_1'\theta_0) x_1 \right\|^2 \\
&\leq O(1) \left\{ E\left| \delta_{0,k}^{(1)}(x_1'\theta_0) \right|^4 E\|x_1\|^4 \right\}^{1/2} = O(1) \left(E\left| \delta_{0,k}^{(1)}(x_1'\theta_0) \right|^4 \right)^{1/2} = o(1).
\end{aligned}$$

Therefore, $J_{3N} = o_P(1)$.

For J_{5N}, write

$$\begin{aligned}
&\left\| \frac{1}{N} \sum_{i=1}^{N} \delta_{0,k}(x_i'\theta_0) \widetilde{g}_k^{(1)}(x_i'\theta_0) x_i \right\| \leq \frac{1}{N} \sum_{i=1}^{N} \left\| \delta_{0,k}(x_i'\theta_0) \widetilde{g}_k^{(1)}(x_i'\theta_0) x_i \right\| \\
&\leq \left(\frac{1}{N} \sum_{i=1}^{N} \|\delta_{0,k}(x_i'\theta_0)\|^2 \right)^{1/2} \left(\frac{1}{N} \sum_{i=1}^{N} \left\| \widetilde{g}_k^{(1)}(x_i'\theta_0) x_i \right\|^2 \right)^{1/2} = O_P(k^{-r/2}),
\end{aligned}$$

where the last equality follows from the facts that $E\left[\frac{1}{N}\sum_{i=1}^{N} \|\delta_{0,k}(x_i'\theta_0)\|^2\right] = O(k^{-r})$ by Assumption 1.2 and Lemma A.1, and $E\left[\frac{1}{N}\sum_{i=1}^{N} \left\|\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i\right\|^2\right] = O(1)$ by Assumption 1.2. Thus, $J_{5N} = O_P\left(\sqrt{Nk^{-r}}\right) = o_P(1)$.

For J_{7N}, write

$$\begin{aligned}
&\frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \widetilde{g}_k^{(1)}(x_i'\widetilde{\theta}) x_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \left[\widetilde{g}_k^{(1)}(x_i'\theta_0) + \widetilde{g}_k^{(1)}(x_i'\widetilde{\theta}) - \widetilde{g}_k^{(1)}(x_i'\theta_0) \right] x_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \left[\widetilde{g}_k^{(2)}(x_i'\theta^*)(\widetilde{\theta} - \theta_0)' x_i \right] x_i \\
&\quad + \frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \widetilde{g}_k^{(1)}(x_i'\theta_0) x_i \\
&= \frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \widetilde{g}_k^{(2)}(x_i'\theta^*) x_i x_i' (\widetilde{\theta} - \theta_0) \\
&\quad + \frac{1}{N} \sum_{i=1}^{N} \left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta}) \right] \widetilde{g}_k^{(1)}(x_i'\theta_0) x_i
\end{aligned}$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\theta_0) + \widetilde{g}_k(x_i'\theta_0) - \widetilde{g}_k(x_i'\widetilde{\theta})\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i + o_P(\|\theta_0 - \widetilde{\theta}\|)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\left[\widetilde{g}_k^{(1)}(x_i'\theta_1^*)(\theta_0 - \widetilde{\theta})'x_i\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i + o_P(\|\theta_0 - \widetilde{\theta}\|)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\widetilde{g}_k^{(1)}(x_i'\theta_1^*)\widetilde{g}_k^{(1)}(x_i'\theta_0)x_ix_i'(\theta_0 - \widetilde{\theta}) + o_P(\|\theta_0 - \widetilde{\theta}\|), \tag{A.10}$$

where θ^* and θ_1^* both lie between $\widetilde{\theta}$ and θ_0; the second and fifth equalities follow from Mean Value Theorem; the the fourth equality follows from the proof similar to (1) of this lemma. Therefore, we can write

$$\theta_0\widetilde{\theta}'J_{7N} - J_{4N} = (-I + \theta_0\widetilde{\theta}')S_N(\widetilde{g}_k) + \theta_0\widetilde{\theta}'T_N(\widetilde{g}_k)(\theta_0 - \widetilde{\theta}), \tag{A.11}$$

where

$$S_N(\widetilde{g}_k) = \frac{1}{N}\sum_{i=1}^{N}\left[g_0(x_i'\theta_0) - \widetilde{g}_k(x_i'\theta_0)\right]\widetilde{g}_k^{(1)}(x_i'\theta_0)x_i,$$

$$T_N(\widetilde{g}_k) = \frac{1}{N}\sum_{i=1}^{N}\widetilde{g}_k^{(1)}(x_i'\theta_1^*)\widetilde{g}_k^{(1)}(x_i'\theta_0)x_ix_i'.$$

Notice that the matrix $I - \theta_0\widetilde{\theta}$ in front of J_{1N} converges to the projection matrix $P_{\theta_0} = I_d - \theta_0\theta_0'$ in probability, which has eigenvalues $0, 1, \ldots, 1$ and 0 has eigenvector θ_0. Therefore, in order to ensure the asymptotic covariance matrix is non-singular, we need to rotate $\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta}$. Let $P_1 = (p_1, \ldots, p_{d-1})$, where $p_1, \ldots, p_{d-1}$ are the eigenvectors associated with the non-zero eigenvalues of P_{θ_0}. Thus, we have $P_{\theta_0} = P_1P_1'$ and $P_1'P_1 = I_{d-1}$. Having said that, we then obtain that

$$\sqrt{N}P_1'\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})} + \sqrt{N}2P_1'\theta_0\widetilde{\theta}'T_N(\widetilde{g}_k) + P_1'P_{\theta_0}\sqrt{N}2S_N(\widetilde{g}_k)$$

$$= \sqrt{N}P_1'\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})} + P_1'\sqrt{N}2S_N(\widetilde{g}_k) \to_D N(0, 4\sigma_e^2P_1'VP_1),$$

where the first equality follows from the fact that $P_1'\theta_0 = 0$.

The proof is then completed. ■

Proof of Theorem 2.2:

(1). Notice that

$$\widetilde{\theta}'\theta_0 - 1 = (\widetilde{\theta} - \theta_0)'\theta_0 = (\widetilde{\theta} - \theta_0)'(\theta_0 - \widetilde{\theta} + \widetilde{\theta})$$

$$= -\|\widetilde{\theta} - \theta_0\|^2 + (\widetilde{\theta} - \theta_0)'\widetilde{\theta} = -\|\widetilde{\theta} - \theta_0\|^2 + (1 - \widetilde{\theta}'\theta_0). \tag{A.12}$$

Thus, we have $1 - \widetilde{\theta}'\theta_0 = \frac{1}{2}\|\widetilde{\theta} - \theta_0\|^2$.

(2). Following the routine procedure documented in the literature (c.f. Yu and Ruppert (2002)), write

$$0 = P_1'\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta}\Big|_{(\theta,C_k,\lambda)=(\widetilde{\theta},\widetilde{C}_k,\widetilde{\lambda})}$$

$$= P_1'\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})} + P_1'\frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'}\Big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})}\left(\widetilde{\theta} - \theta_0\right)$$

$$
\begin{aligned}
&= P_1' \frac{\partial \mathcal{W}_{N,\lambda}}{\partial \theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})} + P_1' \frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'}\Big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})} (P_1P_1' + \theta_0\theta_0') \left(\widetilde{\theta} - \theta_0\right) \\
&= P_1' \frac{\partial \mathcal{W}_{N,\lambda}}{\partial \theta}\Big|_{(\theta,C_k,\lambda)=(\theta_0,\widetilde{C}_k,\widetilde{\lambda})} + P_1' \frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'} P_1\Big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})} P_1' \left(\widetilde{\theta} - \theta_0\right) \\
&\quad - \frac{1}{2}\|\widetilde{\theta} - \theta_0\|^2 P_1' \frac{\partial \mathcal{W}_{N,\lambda}}{\partial\theta\partial\theta'}\Big|_{(\theta,C_k,\lambda)=(\overline{\theta},\widetilde{C}_k,\widetilde{\lambda})} \theta_0, \qquad \text{(A.13)}
\end{aligned}
$$

where $\overline{\theta}$ lies between θ_0 and $\widetilde{\theta}$; the third equality follows from $I_d = P_{\theta_0} + \theta_0\theta_0' = P_1P_1' + \theta_0\theta_0'$. Following from (2.11), Lemma A.2 and (1) of this theorem, we immediately obtain that

$$
\sqrt{N}\left(P_1'(\widetilde{\theta} - \theta_0) - 2(P_1'VP_1)^{-1}P_1'S_N(\widetilde{g}_k)\right) \to_D N(0, \sigma_e^2(P_1'VP_1)^{-1}).
$$

Then the proof is complete. ■

Lemma A.3. *Suppose that $u = (u_1, \ldots, u_d)'$, $\nu = (\nu_1, \ldots, \nu_d)' \in \mathbb{R}^d$ and $\|\nu\| = 1$. Then, for $H_m(\cdot)$ defined in (2.1), we have $H_m(u'\nu) = \sum_{|p|=m} \binom{m}{p} \prod_{j=1}^d H_{p_j}(u_j) \prod_{j=1}^d \nu_j^{p_j}$, where $p = (p_1, \ldots, p_d)$, p_j for $j = 1, \ldots, d$ are all non-negative integers, $|p| = p_1 + \cdots + p_d$ and $\binom{m}{p} = \frac{m!}{\prod_{j=1}^d p_j!}$.*

The detailed proof of this lemma has been given in Dong et al. (2015), thus omitted.

For notational simplicity, we denote some variables before proceed further. For $\theta \in \Theta$,

$$
\begin{aligned}
&\underset{N\times k}{\mathsf{H}_N}(\theta) = (\mathcal{H}(x_1'\theta), \ldots, \mathcal{H}(x_N'\theta)), \quad \underset{N\times 1}{\mathcal{G}_N}(\theta) = (g_0(x_1'\theta), \ldots, g_0(x_N'\theta))', \\
&\underset{N\times 1}{\mathcal{D}_N}(\theta) = (\delta_{0,k}(x_1'\theta), \ldots, \delta_{0,k}(x_N'\theta))', \quad \underset{N\times 1}{\mathcal{E}_N} = (e_1, \ldots, e_N)', \\
&\underset{N\times K}{\mathcal{Z}_N} = (\mathcal{Z}(x_1), \ldots, \mathcal{Z}(x_N))'. \qquad \text{(A.14)}
\end{aligned}
$$

Lemma A.4. *Under Assumptions 1, 3 and 4, we have*

1. $\left\|\frac{1}{N}\mathcal{Z}_N'\mathcal{E}_N\right\| = O_P\left(\sqrt{\frac{k^d}{N}}\right)$;
2. $\left\|\frac{1}{N}\mathcal{Z}_N\mathcal{Z}_N' - \mathcal{Z}_x\right\| = o_P(1)$, *where* $\mathcal{Z}_x = E[\mathcal{Z}(x_1)\mathcal{Z}(x_1)']$;
3. $\sup_\theta \left\|\frac{1}{N}\mathsf{H}_N(\theta)'\mathcal{E}_N\right\| = O_P\left(\sqrt{\frac{k}{N}}\right)$;
4. $\sup_\theta \left\|\frac{1}{N}\mathsf{H}_N(\theta)\mathsf{H}_N(\theta)' - \Sigma(\theta)\right\| = o_P(1)$, *where ϵ is a sufficiently small number.*

Proof of Lemma A.4:

(1) Write

$$
\begin{aligned}
E\left\|\frac{1}{N}\mathcal{Z}_N'\mathcal{E}_N\right\|^2 &= \frac{\sigma_e^2}{N^2}\sum_{m=1}^K\sum_{i=1}^N E\left|z_m(x_i)\right|^2 = \frac{\sigma_e^2}{N}\sum_{m=1}^K \int_{\mathbb{R}^d} |\mathscr{H}_{\mathsf{p}_m}(z)|^2 \cdot f(z)dz \\
&= \frac{\sigma_e^2}{N}\sum_{m=1}^K \int_{\mathbb{R}^d} |\mathscr{H}_{\mathsf{p}_m}(z)|^2 \exp(-\|z\|^2/2) \cdot \exp(\|z\|^2/2) f(z)dz \\
&\le O(1)\frac{1}{N}\sum_{m=1}^K \int_{\mathbb{R}^d} |\mathscr{H}_{\mathsf{p}_m}(z)|^2 \exp(-\|z\|^2/2)dz \\
&= \frac{1}{N}\sum_{m=0}^{k-1}\sum_{|\mathsf{p}|=m}\left(\prod_{j=1}^d \int_{\mathbb{R}} h_{p_j}^2(w)\exp(-w^2/2)dw\right) = O\left(\frac{K}{N}\right), \qquad \text{(A.15)}
\end{aligned}
$$

where $f(\cdot)$ denotes the pdf of x, and the fist inequality follows from Assumption 3.1.

(2) Write

$$E\left\|\frac{1}{N}\mathcal{Z}_N\mathcal{Z}_N' - \mathcal{Z}_x\right\|^2 = \frac{1}{N}\sum_{m=1}^{K}\sum_{n=1}^{K}E|z_m(x)z_n(x) - \mathcal{Z}_{x,mn}|^2$$
$$\leq \frac{1}{N}\sum_{m=1}^{K}\sum_{n=1}^{K}\int_{\mathbb{R}^d}|\mathscr{H}_{\mathsf{p}_m}(z)|^2|\mathscr{H}_{\mathsf{p}_n}(z)|^2 f(z)dz = \Psi(N,k) = o(1), \tag{A.16}$$

where p_m with $m = 1,\ldots,K$ is defined in (2.17); $\mathcal{Z}_{x,mn}$ denotes the $(m,n)^{th}$ element of $\mathcal{Z}_x$; the last equality follows from Assumption 3.3.

(3) We now prove this by Lemma A.2 of Newey and Powell (2003). By the same spirit of (B.11) of Chen et al. (2012), it suffices to show that

$$\sup_\theta\left\|\frac{1}{N}\mathsf{H}_N(\theta)'\mathcal{E}_N\right\| = o_P\left(l(N)\sqrt{\frac{k}{N}}\right),$$

where $l(\cdot)$ is any positive function that satisfies $l(N)\to\infty$ as $N\to\infty$.

Again, we use Lemma A2 of Newey and Powell (2003). Similar to the proof of Theorem 2.1, we just need to show that for $\forall\theta$, $\frac{N}{l(N)k}E\left\|\frac{1}{N}\mathsf{H}_N(\theta)'\mathcal{E}_N\right\|^2 = o(1)$.

Note that

$$E\left\|\frac{1}{N}\mathsf{H}_N(\theta)'\mathcal{E}_N\right\|^2 = \frac{\sigma_e^2}{N^2}\sum_{i=1}^{N}E\left\|\mathcal{H}(x_i'\theta)\right\|^2 = \frac{\sigma_e^2}{N}E\left\|\mathcal{H}(x'\theta)\right\|^2$$
$$= \frac{\sigma_e^2}{N}\sum_{j=0}^{k-1}\int h_u^2(w)f_\theta(w)dw = \frac{\sigma_e^2}{N}\sum_{j=0}^{k-1}\int h_u^2(w)\exp(-w^2/2)\cdot\exp(w^2/2)f_\theta(w)dw$$
$$\leq O(1)\frac{1}{N}\sum_{j=0}^{k-1}\int h_u^2(w)\exp(-w^2/2)dw = O\left(\frac{k}{N}\right),$$

where $f_\theta(\cdot)$ is denoted in Assumption 1.2, and the first inequality follows from Assumption 1.2. Thus, we have that $\left\|\frac{1}{N}\mathsf{H}_N(\theta)'\mathcal{E}_N\right\| = o_P\left(l(N)\sqrt{\frac{k}{N}}\right)$ for $\forall\theta\in\Theta$. In connection with Lemma A2 of Newey and Powell (2003), the result follows immediately.

(4) By Assumption 1.1, we have

$$E\left\|\frac{1}{N}\mathsf{H}_N(\theta)\mathsf{H}_N(\theta)' - \Sigma(\theta)\right\|^2 = \frac{1}{N}\sum_{m=1}^{k}\sum_{n=1}^{k}E|h_m(x'\theta)h_n(x'\theta) - \Sigma_{mn}(\theta)|^2$$
$$\leq \frac{1}{N}\sum_{m=1}^{k}\sum_{n=1}^{k}\int_{\mathbb{R}^d}h_m^2(w)\,h_n^2(w)\,f_\theta(w)dw = \Phi(N,k) = o(1), \tag{A.17}$$

where $\Sigma_{mn}(\theta)$ denotes the $(m,n)^{th}$ element of $\Sigma(\theta)$; the last equality follows from Assumption 4. Similar to result (3) of this lemma, the result follows from Lemma A2 of Newey and Powell (2003). ■

Proof of Lemma 2.1:

Expanding (2.19), we obtain

$$\widehat{\beta}_K - \beta_{0,K} = \left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)\mathcal{Z}(x_i)'\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)e_i$$
$$+\left(\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)\mathcal{Z}(x_i)'\right)^{-1}\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)\delta_{0,k}(x_i'\theta_0).$$

By Lemma A.4, we already have

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)e_i\right\| = O_P\left(\sqrt{\frac{k^d}{N}}\right) \quad\text{and}\quad \left\|\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)\mathcal{Z}(x_i)' - \mathcal{Z}_x\right\| = o_P(1).$$

Then focus on $\frac{1}{N}\sum_{i=1}^{N}\mathcal{Z}(x_i)\,\delta_{0,k}(x_i'\theta_0)$. Write

$$\begin{aligned}\left\|(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'\mathcal{D}_N(\theta_0)\right\|^2 &= \mathcal{D}_N(\theta_0)'\mathcal{Z}_N(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}(\mathcal{Z}_N'\mathcal{Z}_N/N)^{-1}\mathcal{Z}_N'\mathcal{D}_N(\theta_0)/N\\ &\le \lambda_{min}^{-1}(\mathcal{Z}_N'\mathcal{Z}_N/N)\cdot\mathcal{D}_N(\theta_0)'\mathcal{Z}_N(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'\mathcal{D}_N(\theta_0)/N\\ &\le \lambda_{min}^{-1}(\mathcal{Z}_N'\mathcal{Z}_N/N)\cdot\lambda_{max}(W)\cdot\|\mathcal{D}_N(\theta_0)\|^2/N, \end{aligned}\tag{A.18}$$

where the first inequality follows from Magnus and Neudecker (2007, exercise 5 on p. 267). Note that $W=\mathcal{Z}_N(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'$ is symmetric and idempotent, so $\lambda_{max}(W)=1$. By Assumption 1.1, it is easy to know that $E[\|\mathcal{D}_N(\theta_0)\|^2/N]=E|\delta_{0,k}(x'\theta_0)|^2=O(k^{-r})$, where the last equality follows from Lemma A.1. Therefore, $\|\mathcal{D}_N(\theta_0)\|^2/N=O_P(k^{-r})$. Moreover, by (2) of Lemma A.4 and Assumption 3.2, we know that $\lambda_{min}^{-1}(\mathcal{Z}_N'\mathcal{Z}_N/N)=O_P(1)$. Thus, we obtain $\left\|(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'\mathcal{D}_N(\theta_0)\right\|=O_P(k^{-r/2})$.

According to the above derivations, we obtain $\|\widehat{\beta}_K-\beta_{0,K}\|=O_P\left(\sqrt{\frac{k^d}{N}}\right)+O_P(k^{-r/2})$. The proof is then complete. ■

Proof of Theorem 2.3:

Notice that we have denoted $\mathcal{Z}_N$, $\mathcal{D}_N(\theta_0)$ and $\mathcal{E}_N$ in (A.14).

$$\begin{aligned}\widehat{\theta}-\theta_0 &= \frac{1}{\widehat{c}_1}(Q\widehat{\beta}-Q\beta)+\frac{1}{\widehat{c}_1}(c_1-\widehat{c}_1)\theta_0\\ &=\frac{1}{\widehat{c}_1}\left\{(Q\widehat{\beta}-Q\beta)-\theta_0(\widehat{c}_1-c_1)\right\}\\ &=\frac{1}{\widehat{c}_1}\left\{(Q\widehat{\beta}-Q\beta)-\frac{1}{\widehat{c}_1+c_1}\theta_0(\widehat{c}_1^2-c_1^2)\right\}\\ &=\frac{1}{\widehat{c}_1}\left\{(Q\widehat{\beta}-Q\beta)-\frac{1}{\widehat{c}_1+c_1}\theta_0\left[(Q\widehat{\beta})'(Q\widehat{\beta})-(Q\beta)'(Q\beta)\right]\right\}\\ &=\frac{1}{\widehat{c}_1}\left\{(Q\widehat{\beta}-Q\beta)-\frac{1}{\widehat{c}_1+c_1}\theta_0\left[Q\widehat{\beta}-Q\beta\right]'\left[Q\widehat{\beta}+Q\beta\right]\right\}\\ &=\frac{1}{\widehat{c}_1}\left\{(Q\widehat{\beta}-Q\beta)-\frac{1}{\widehat{c}_1+c_1}\left[\left(Q\widehat{\beta}+Q\beta\right)'\otimes\theta_0\right]\left[Q\widehat{\beta}-Q\beta\right]\right\}\\ &=\frac{1}{\widehat{c}_1}\left\{I_d-\frac{1}{\widehat{c}_1+c_1}\left[\left(Q\widehat{\beta}+Q\beta\right)'\otimes\theta_0\right]\right\}(Q\widehat{\beta}-Q\beta),\end{aligned}\tag{A.19}$$

where the third equality follows from $\widehat{c}_1^2-c_1^2=(\widehat{c}_1-c_1)(\widehat{c}_1+c_1)$; the fourth equality follows from the definition of $\widehat{c}_1$ and $\|\theta_0\|=1$; the sixth equality follows from $\text{vec}(ABC)=(C'\otimes A)\text{vec}(B)$.

Notice that $I_d-\frac{1}{\widehat{c}_1+c_1}\left[\left(Q\widehat{\beta}+Q\beta\right)'\otimes\theta_0\right]$ in the asymptotic covariance, which is

$$\begin{aligned}&I_d-\frac{1}{\widehat{c}_1+c_1}\left[\left(Q\widehat{\beta}+Q\beta\right)'\otimes\theta_0\right]\\ \to_P\ &I_d-\frac{1}{2c_1}\left[2c_1\theta_0'\otimes\theta_0\right]=I_d-\theta_0'\otimes\theta_0=I_d-\theta_0\theta_0'=M_{\theta_0}=P_1P_1',\end{aligned}\tag{A.20}$$

where P_1 has been defined.

Therefore, in order to establish the asymptotic normality, we need to consider $P_1'(\widehat{\theta}-\theta_0)$. Below we focus on $\sqrt{N}Q\left(\widehat{\beta}_K-\beta_{0,K}\right)$ and write

$$\sqrt{N}Q\left(\widehat{\beta}_K-\beta_{0,K}\right)=\sqrt{N}Q(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'\mathcal{D}_N(\theta_0)+\sqrt{N}Q(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'\mathcal{E}_N.\tag{A.21}$$

Note that $K=O\left(k^d\right)$ and $Q=O(1)$. In connection with the proof of Lemma (2.1), it is straightforward to obtain

$$\left\|\sqrt{N}Q(\mathcal{Z}_N'\mathcal{Z}_N)^{-1}\mathcal{Z}_N'\mathcal{D}_N(\theta_0)\right\|=O_P\left(N^{1/2}k^{-r/2}\right)=o_P(1),$$

where the last equality follows from the condition in the body of this theorem.

Then, in order to esatblish the normality, we need only to consider the second term on RHS of (A.21):

$$\sqrt{N}Q\left(\mathcal{Z}_N'\mathcal{Z}_N\right)^{-1}\mathcal{Z}_N'\mathcal{E}_N = Q\mathcal{Z}_x^{-1}\frac{1}{\sqrt{N}}\mathcal{Z}_N'\mathcal{E}_N(1+o_P(1))$$
$$= Q\mathcal{Z}_x^{-1}\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathcal{Z}(x_i)e_i(1+o_P(1)) \to_D N(0,\sigma_e^2\Omega_1),$$

where $\Omega_1 = \lim_{k\to\infty}\sigma_e^2 Q\mathcal{Z}_x^{-1}Q'$, the first equality follows from (2) of Lemma A.4, and the last line follows from Assumption 1.1 and CLT. In connection with (A.19) and (A.20), the proof is then completed. ■

Proof of Theorem 2.4:

The proof of this theorem follows from exactly the same as that used for Theorem 2.3. ■

Proof of Theorem 2.5:

(1). Write

$$\widehat{C}_k - C_{0,k} = \left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1}\mathsf{H}_N(\widehat{\theta})'\left(\mathcal{G}_N(\theta_0)-\mathcal{G}_N(\widehat{\theta})\right)$$
$$+\left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1}\mathsf{H}_N(\widehat{\theta})'\mathcal{D}_N(\widehat{\theta})+\left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1}\mathsf{H}_N(\widehat{\theta})'\mathcal{E}_N, \tag{A.22}$$

where $\mathcal{G}_N(\theta)$, $\mathsf{H}_N(\theta)$, $\mathcal{E}_N$ and $\mathcal{D}_N(\theta)$ have been defined in (A.14) respectively.

By the same procedure as (A.18), it is easy to know that

$$\left\|\left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1}\mathsf{H}_N(\widehat{\theta})'\mathcal{D}_N(\widehat{\theta})\right\| = O_P\left(k^{-r/2}\right).$$

By (4) of Lemma A.4 and Assumption 2.3, we have

$$\left\|\left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1}\mathsf{H}_N(\widehat{\theta})'\mathcal{E}_N\right\| = O_P\left(\frac{k^{1/2}}{N^{1/2}}\right).$$

Then, we need only to consider the next term. Write

$$\left\|\left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1}\mathsf{H}_N(\widehat{\theta})'\left(\mathcal{G}(\theta_0)-\mathcal{G}(\widehat{\theta})\right)\right\|^2$$
$$\leq \left(\lambda_{min}\left(\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})/N\right)\right)^{-1}\cdot\lambda_{max}(\widetilde{W})\cdot\left(\left\|\mathcal{G}(\theta_0)-\mathcal{G}(\widehat{\theta})\right\|^2/N\right),$$

where $\widetilde{W} = \mathsf{H}_N(\widehat{\theta})\left(\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right)^{-1}\mathsf{H}_N(\widehat{\theta})'$. Notice that for $\forall\theta\in\{\theta:\|\theta-\theta_0\|<\epsilon\}$, we have

$$\frac{1}{N}E\left[\|\mathcal{G}(\theta_0)-\mathcal{G}(\theta)\|^2\right] = \frac{1}{N}\sum_{i=1}^{N}E\left[g_0(x_i'\theta_0)-g_0(x_i'\theta)\right]^2$$
$$= E\left[\left(g_0^{(1)}(x_1'\theta^*)\right)^2(\theta_0-\theta)'x_1x_1'(\theta_0-\theta)\right] = \|\theta_0-\theta\|^2 E\left\|\left(g_0^{(1)}(x_1'\theta_*)\right)^2 x_1x_1'\right\|$$
$$= O(1)\|\theta_0-\theta\|^2,$$

where θ^* lies between θ_0 and θ, and the second equality follows from Assumption 1.1. Therefore, we have

$$\frac{1}{N}\left\|\mathcal{G}(\theta_0)-\mathcal{G}(\widehat{\theta})\right\|^2 = O_P(1)\|\theta_0-\widehat{\theta}\|^2.$$

Since $\widetilde{W}$ is symmetric and idempotent, $\lambda_{max}(\widetilde{W}) = 1$. Hence, we have

$$\left\| \left[\mathsf{H}_N(\widehat{\theta})'\mathsf{H}_N(\widehat{\theta})\right]^{-1} \mathsf{H}_N(\widehat{\theta})' \left(\mathcal{G}(\theta_0) - \mathcal{G}(\widehat{\theta})\right) \right\| = O_P(1)\|\theta_0 - \theta\| = O_P\left(\frac{1}{\sqrt{N}}\right).$$

Based on the above derivations, the result follows.

(2). By the orthogonality, we write

$$\begin{aligned}
\|\widehat{g}_k(w) - g_0(w)\|_{L^2}^2 =& \int_{\mathbb{R}} \left(\widehat{g}_k(w) - g_0(w)\right)^2 \cdot \exp\left(-w^2/2\right) dw \\
=& \int_{\mathbb{R}} \left[\mathcal{H}(w)'(\widehat{C}_k - C_{0,k}) + \delta_{0,k}(w)\right]^2 \cdot \exp\left(-w^2/2\right) dw \\
=& O\left(\|\widehat{C}_k - C_{0,k}\|^2\right) + O\left(\|\delta_{0,k}(w)\|_{L^2}^2\right) \\
=& O_P\left(\frac{k}{N}\right) + O_P\left(k^{-r}\right) + O\left(k^{-r}\right) = O_P\left(\frac{k}{N}\right) + O_P\left(k^{-r}\right),
\end{aligned}$$

where the fourth equality follows from the first result of this theorem and Lemma A.1. Then the proof is completed. ■

References

Ai, C. and Chen, X. (2003), 'Efficient estimation of models with conditional moment restrictions containing unknown functions', *Econometrica* **71**(6), 1795–1843.

Amemiya, T. (1993), 'Non-linear regression model', *Handbook of Econometrics* pp. 333–389.

Belloni, A., Chernozhukov, V., Chetverikov, D. and Kato, K. (2015), 'Some new asymptotic theory for least squares series: Pointwise and uniform results', *Journal of Econometrics* **186**, 345–366.

Berger, A. and Mester, L. (2003), 'Explaining the dramatic changes in performance of U.S. banks: technical change, deregulation, and dynamic changes in competition', *Journal of Financial Intermediation* **12**(1), 57–95.

Birke, M., Bellegem, S. and Keilegom, I. (2017), 'Semiparametric estimation in a single-index model with endogenous variable', *Scandinavian Journal of Statistics* **44**, 168–191.

Cai, Z., Juhl, T. and Yang, B. (2015), 'Functional index coefficient models with variable selection', *Journal of Econometrics* **189**, 272–284.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997), 'Generalized partially linear single-index models', *Journal of the American Statistical Association* **92**(438), 477–489.

Chang, J., Chen, S. and Chen, X. (2015), 'High dimensional generalized empirical likelihood for moment restrictions with dependent data', *Journal of Econometrics* **185**, 283–304.

Chen, J., Gao, J. and Li, D. (2012), 'Semiparametric trending panel data models with cross-sectional dependence', *Journal of Econometrics* **171**(1), 71–85.

Chen, X. (2007), 'Large sample sieve estimation of semi-nonparametric models', *Handbook of Econometrics* pp. 5549–5632.

Chen, X. and Christensen, T. (2015), 'Optimal uniform convergence rates and asymptotic normality for series estimators under weak dependence and weak conditions', *Journal of Econometrics* **188**, 447–465.

Chen, X., Linton, O. and Keilegom, I. V. (2003), 'Estimation of semiparametric models when the criterion function is not smooth', *Econometrica* **71**(5), 1591–1608.

Chu, W., Li, R. and Reimherr, M. (2016), 'Feature screening for time-varying coefficient models with ultra–high dimensional longitudinal data', *Annals of Applied Statistics* **10**(2), 596–617.

Dong, C., Gao, J. and Peng, B. (2015), 'Semiparametric single-index panel data model with cross-sectional dependence', *Journal of Econometrics* **188**(1), 301–312.

Dong, C., Gao, J. and Tjøstheim, D. (2016), 'Estimation for single-index and partially linear single-index nonstationary time series models', *Annals of Statistics* **44**(1), 425–453.

Feng, G. and Serletis, A. (2008), 'Productivity trends in U.S. manufacturing: Evidence from the nq and aim cost functions', *Journal of Econometrics* **142**(1), 281–311.

Feng, G. and Zhang, X. (2012), 'Productivity and efficiency at large and community banks in the U.S.: a Bayesian true random effects stochastic distance frontier analysis', *Journal of Banking and Finance* **36**(7), 1883–1895.

Gao, J. (2007), *Nonlinear Time Series: Semi– and Non–Parametric Methods*, Chapman & Hall/CRC.

Gao, J., Tong, H. and Wolff, R. (2002), 'Adaptive series estimation in additive stochastic regression models', *Statistica Sinica* **12**, 409–428.

Hansen, X. (2015), 'A unified asymptotic distribution theory for parametric and non-parametric least squares', *Working paper* .

Härdle, W., Hall, P. and Ichimura, H. (1993), 'Optimal smoothing in single-index models', *Annals of Statistics* **21**(1), 157–178.

Henderson, D. J., Kumbhakar, S. C., Li, Q. and Parmeter, C. F. (2015), 'Smooth coefficient estimation of a seemingly unrelated regression', *Journal of Econometrics* **189**(1), 148–162.

Ichimura, H. (1993), 'Semiparametric least squares (SLS) and weighted SLS estimation of single-index models', *Journal of Econometrics* **59**(1-2), 71–120.

Ma, S. and Song, P. X. K. (2015), 'Varying index coefficient models', *Journal of the American Statistical Association* **110**(509), 341–356.

Magnus, J. R. and Neudecker, H. (2007), *Matrix Differential Calculus with Applications in Statistics and Econometrics*, third edn, John Wiley & Sons Ltd.

Newey, W. K. (1994), 'The asymptotic variance of semiparametric estimators', *Econometrica* **62**(6), 1349–1382.

Newey, W. K. (1997), 'Convergence rates and asymptotic normality for series estimators', *Journal of Econometrics* pp. 147–168.

Newey, W. K. and Powell, J. L. (2003), 'Instrumental variable estimation of nonparametric models', *Econometrica* **71**(5), 1565–1578.

Sealey, C. W. and Lindley, J. T. (1977), 'Inputs, outputs, and a theory of production and cost at depository financial institutions', *Journal of Finance* **32**(4), 1251–1266.

Su, L. and Jin, S. (2012), 'Sieve estimation of panel data models with cross section dependence', *Journal of Econometrics* **169**(1), 34–47.

Xia, Y. (2006), 'Asymptotic distribution for two estimators of the single-index model', *Econometric Theory* **22**(6), 1112–1137.

Xia, Y., Tong, H. and Li, W. K. (1999), 'On extended partially linear single-index models', *Boimetrika* **86**(4), 831–842.

Xia, Y., Tong, H., Li, W. K. and Zhu, L.-X. (2002), 'An adaptive estimation of dimension reduction', *Journal of the Royal Statistical Society B* **64**(3), 363–410.

Yu, Y. and Ruppert, D. (2002), 'Penalized spline estimation for partially linear single-index models', *Journal of the American Statistical Association* **97**(460), 1042–1054.