

Semiparametric Single–Index Panel Data Models with Cross–Sectional Dependence

CHAOHUA DONG^{*,*}, JITI GAO^{*} AND BIN PENG^{†1}

^{*}Monash University and [†]University of Technology Sydney
and ^{*}Southwestern University of Finance and Economics

Abstract

In this paper, we consider a semiparametric single–index panel data model with cross–sectional dependence and stationarity. Meanwhile, we allow fixed effects to be correlated with the regressors to capture unobservable heterogeneity. Under a general spatial error dependence structure, we then establish some consistent closed–form estimates for both the unknown parameters and the link function for the case where both cross–sectional dimension (N) and temporal dimension (T) go to infinity. Rates of convergence and asymptotic normality are established for the proposed estimates. Our experience suggests that the proposed estimation method is simple and thus attractive for finite–sample studies and empirical implementations. Moreover, both the finite–sample performance and the empirical applications show that the proposed estimation method works well when the cross–sectional dependence exists in the data set.

Keywords: Asymptotic theory; closed–form estimate; nonlinear panel data model; orthogonal series expansion method.

JEL classification: C13, C14, C23, C51

¹*Address for Correspondence:* Bin Peng, Economics Discipline Group, University of Technology Sydney, Australia. Email: Bin.Peng@uts.edu.au.

1 Introduction

Single-index models have been studied by both econometricians and statisticians in the past twenty years or so and cover many classic parametric models (e.g. linear model and logistic model) by using a general function form $g(x'\beta)$ (e.g. Chapter 2 of Gao (2007)). Recently, researchers start considering single-index models under panel data set-up (c.f. Chen et al. (2013a) and Chen et al. (2013b)). For most of the published work on semiparametric single-index models, the estimation is based on a nonparametric kernel method, which may be sensitive to initial values due to multi-modality or flatness of a curve in practice. Chen et al. (2013b) use this technique to investigate a partially linear panel data model with fixed effects and cross-sectional independence. In their paper, a consistent parameter estimator is achieved with the rate of convergence \sqrt{NT} (N and T here and hereafter are cross-sectional dimension and temporal dimension, respectively), but, due to the identification requirements, they have to impose extra restrictions on the fixed effects. Alternatively, one can use sieve estimation techniques to implement a two-step procedure (e.g. profile method or iterative method).

To the best of our knowledge, consistent closed-form estimates have not been established for this type of semiparametric single-index model in the literature. In this paper, we aim at establishing consistent closed-form estimates for a semiparametric single-index panel data model with both cross-sectional dependence and stationarity for the case where both N and T go to infinity. The estimation procedure proposed below allows us to avoid certain computational issues and is therefore easy to implement. The estimation techniques proposed in this paper can also be extended to the multi-factor structure model. For example, under certain restrictions similar to those in Su and Jin (2012), a semiparametric single-index extension of Pesaran (2006) can be achieved. Furthermore, we add fixed effects to the model and do not impose any particular assumptions on them, and therefore they can be correlated with the regressors to capture unobservable heterogeneity. Compared to Chen et al. (2013b), our set-up is more flexible on the fixed effects.

In this paper, we assume that all the regressors and error terms can be cross-sectionally correlated. As covered in Assumption 1 of Section 3 below, we impose a general spatial correlation structure to link the cross-sectional dependence and stationary mixing condition together. As a result, some types of spatial error correlation can easily be covered by the assumptions given in Section 3 (c.f. Chen et al. (2012a) and Chen et al. (2012b)). This set-up is more flexible than that considered by Chen et al. (2013b). As Andrews (2005) and Bai (2009) discuss, common shocks (e.g. global finance crisis) exist in many economic phenomena and cause serious forecasting biases, and an important characteristic is that they induce a correlation among individuals. Thus, it is vital for us to have such models that can capture this type of “global” cross-sectional dependence.

We specifically use Hermite polynomial series orthogonal expansion to estimate the unknown link function. This technique has been widely used in the econometrical literature, for

example, by Gallant and Nychka (1987) and Gallant and Tauchen (1989), among others. In general, this method is referred to as sieve estimation. Very detailed literature reviews on the Hermite polynomials and the sieve estimation can be seen in Nevai (1986) and Chen (2007) respectively. However, a further technique gadget on the Hermite polynomials brings a benefit to the expansion of the link function (see Lemma B.1 in Appendix below). As a result, a semiparametric model is rephrased completely as a parametric model such that we are able to derive a closed-form estimate for the index vector, and because of this the requirement for the parameter space is lowered to the minimum level, comparing with the profile method where usually θ_0 is stipulated to be an inner point of a compact convex set Θ .

In summary, this paper makes the following contributions:

1. It proposes a semiparametric single-index panel data model to simultaneously accommodate cross-sectional dependence, stationarity and unobservable heterogeneity;
2. It establishes simple and consistent closed-form estimates for the unknown index vector, and consequently there is no restriction on the parameter space;
3. It establishes both rates of convergence and asymptotic normality results for the estimates under a general spatial error dependence structure; and
4. It evaluates the proposed estimation method through the use of both simulated and real data examples.

The structure of this paper is as follows. Section 2 proposes our model and discusses the main idea. Section 3 constructs a closed-form estimate for a vector of unknown parameters of interest and introduces assumptions for the establishment of asymptotic consistency and normality results. In Section 4, we recover the unknown link function and evaluate the rate of convergence. In Section 5, we do Monte Carlo experiments which particularly verify whether the fixed effect dependence affects the proposed estimation and provide an empirical case study by looking into the demand of the United States (US) for cigarettes. Section 6 concludes this paper with some comments. The key proofs are given in the Appendix. Some other proofs, verifications and relevant discussions are given in a supplementary document of this paper.

Throughout the paper, we will use the following notation: \otimes denotes the Kronecker product; I_k denotes the identity matrix with dimension k ; i_k signifies the $k \times 1$ vector $(1, \dots, 1)'$; $M_P = I_k - P(P'P)^{-1}P'$ denotes the projection matrix generated by matrix $P = P_{k \times l}$ with full column rank; A^- denotes the Moore-Penrose inverse of the matrix A ; \rightarrow_P and \rightarrow_D stand for convergence in probability and convergence in distribution, respectively; $\|\cdot\|$ denotes the Euclidean norm; $[a]$ means the largest integer not exceeding a .

2 Semiparametric single-index panel data models

A semiparametric single-index panel data model is specified as follows:

$$y_{it} = g(x'_{it}\theta_0) + \gamma_i + e_{it}, \quad i = 1, \dots, N, \quad t = 1, \dots, T, \quad (2.1)$$

where y_{it} is a scalar process, x_{it} is a $(d \times 1)$ explanatory variable, e_{it} is an error process and the link function $g(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ is unknown. Here, $L^2(\mathbb{R}, \exp(-w^2/2)) = \{g(w) : \int_{\mathbb{R}} g^2(w) \exp(-w^2/2) dw < \infty\}$ is a Hilbert space. We use $\{\gamma_i\}$ to capture fixed effects in this model, which is allowed to be correlated with the regressors. Under the current set-up, our main interest is to consistently estimate the index vector $\theta_0 = (\theta_{01}, \dots, \theta_{0d})'$ and the link function $g(\cdot)$ for the case where both N and T go to infinity.

To meet the identification requirements (c.f. Ichimura (1993) and Horowitz (2009)), we assume that $\|\theta_0\| = 1$ and $\theta_{01} > 0$. For the link function $g(\cdot)$, we expand it by the Hermite polynomials into an orthogonal series and approximate it by a truncated series.

The Hermite polynomial system $\{H_m(w), m = 0, 1, 2, \dots\}$ is a complete orthogonal system in the Hilbert space $L^2(\mathbb{R}, \exp(-w^2/2))$ and each element is defined by

$$H_m(w) = (-1)^m \cdot \exp(w^2/2) \cdot \frac{d^m}{dw^m} \exp(-w^2/2). \quad (2.2)$$

The orthogonality of the system reads $\int_{\mathbb{R}} H_m(w) H_n(w) \exp(-w^2/2) dw = m! \sqrt{2\pi} \delta_{mn}$, where δ_{mn} is the Kronecker delta. Define further that $h_m(w) = \frac{1}{\sqrt{m!}} H_m(w)$, so that for any $g(w) \in L^2(\mathbb{R}, \exp(-w^2/2))$ we can expand it in terms of $h_m(w)$ as follows:

$$g(w) = \sum_{m=0}^{\infty} c_m h_m(w), \quad c_m = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g(w) h_m(w) \exp(-w^2/2) dw. \quad (2.3)$$

Based on the above expansion, one is already able to use a profile method or an iterative estimation method to estimate θ_0 and the link function (e.g. Dong et al. (2014)). Since neither of these two methods results in a closed-form estimator, numerical estimates are often sensitive to the initial values in practice due to multimodality or flatness of a curve. Instead, we further expand $h_m(x'_{it}\theta_0)$ by Lemma B.1 of the Appendix as follows.

$$g(x'_{it}\theta_0) = \sum_{m=0}^{k-1} c_m h_m(x'_{it}\theta_0) + \sum_{m=k}^{\infty} c_m h_m(x'_{it}\theta_0) \quad (2.4)$$

$$= \sum_{m=0}^{k-1} \sum_{|p|=m} a_{mp}(\theta_0) \mathcal{H}_p(x_{it}) + \delta_k(x'_{it}\theta_0), \quad (2.5)$$

where positive integer k is truncation parameter and

$$\begin{aligned}\delta_k(x'_{it}\theta_0) &= \sum_{m=k}^{\infty} c_m h_m(x'_{it}\theta_0), \quad a_{mp}(\theta_0) = \sqrt{\binom{m}{p}} c_m \theta_0^p, \quad \binom{m}{p} = \frac{m!}{\prod_{j=1}^d p_j!}, \\ \theta_0^p &= \prod_{j=1}^d \theta_{0j}^{p_j}, \quad \mathcal{H}_p(x_{it}) = \prod_{j=1}^d h_{p_j}(x_{it,j}), \quad x_{it} = (x_{it,1}, \dots, x_{it,d})', \quad p = (p_1, \dots, p_d)', \\ |p| &= p_1 + \dots + p_d \text{ and } p_j\text{'s for } j = 1, \dots, d \text{ are non-negative integers.}\end{aligned}$$

The expansion (2.5) allows us to separate the covariate x_{it} and the coefficient θ_0 , so the closed-form estimator can be established from it. The term $\delta_k(x'_{it}\theta_0)$ can be considered as a residue after truncation, which goes to zero as k increases (c.f. the supplement of this paper). To further investigate the model, we define an ordering relationship with respect to p in (2.5).

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

Based on Definition 2.1, we list all the $\mathcal{H}_p(x_{it})$'s in the descending order with respect to $|p| = m$ for $m = 0, 1, \dots, k-1$ below.

- As $m = 0$, p in the summation of (2.5) only assumes $p = (0, 0, \dots, 0)'$. Thus,

$$\mathcal{H}_p(x_{it}) \equiv 1, \quad \text{and} \quad a_{0p}(\theta_0) = c_0. \quad (2.6)$$

- As $m = 1$, p in the summation of (2.5) has d possibilities: $q_1 = (1, 0, \dots, 0)'$, \dots , $q_d = (0, \dots, 0, 1)'$. Thus,

$$\mathcal{H}_{q_1}(x_{it}) = x_{it,1}, \quad a_{1q_1}(\theta_0) = c_1 \theta_{01}, \dots, \mathcal{H}_{q_d}(x_{it}) = x_{it,d}, \quad a_{1q_d}(\theta_0) = c_1 \theta_{0d}. \quad (2.7)$$

- As $m = 2, \dots, k-1$, we have $m^* := \binom{m+d-1}{d-1}$ possibilities of vector p :

$$\begin{aligned}q_1 &= (m, 0, \dots, 0)', \quad a_{mq_1}(\theta_0) = c_m \theta_{01}^m, \quad \mathcal{H}_{q_1}(x_{it}) = h_m(x_{it,1}); \\ q_2 &= (m-1, 1, 0, \dots, 0)', \quad a_{mq_2}(\theta_0) = \sqrt{m} c_m \theta_{01}^{m-1} \theta_{02}, \quad \mathcal{H}_{q_2}(x_{it}) = h_{m-1}(x_{it,1}) h_1(x_{it,2}); \\ &\vdots \\ q_{m^*} &= (0, 0, \dots, m)', \quad a_{mq_{m^*}}(\theta_0) = c_m \theta_{0d}^m, \quad \mathcal{H}_{q_{m^*}}(x_{it}) = h_m(x_{it,d}).\end{aligned} \quad (2.8)$$

By (2.6), the first term in (2.5) is the constant c_0 . The second to the $(d+1)^{\text{th}}$ $\mathcal{H}_p(x_{it})$'s in (2.5) are simply x_{it} with coefficients $c_1 \theta_0$ shown in (2.7). In the next section when we recover θ_0 , we shall assume that $c_1 \neq 0$ first, and then we move on to the case where $c_1 = 0, \dots, c_{m_0-1} =$

$0, c_{m_0} \neq 0$, m_0 is finite and fully decided by the link function $g(w)$. Thereby, given k increasing to infinity, we always are able to estimate θ_0 explicitly, as explained below.

By Definition 2.1, let $Z(x_{it}) = (Z_1(x_{it})', \dots, Z_{k-1}(x_{it})')'$ and $\beta = (A_1(\theta_0)', \dots, A_{k-1}(\theta_0'))'$, where, for $m = 1, \dots, k-1$, $Z_m(x_{it})$'s are column vectors consisting of all $\mathcal{H}_p(x_{it})$'s arranged in descent ordering with respect to $p: |p| = m$ and $A_m(\theta_0)$ are column vectors consisting of all corresponding coefficients $a_{mp}(\theta_0)$'s. Therefore, we can rewrite model (2.1) as

$$y_{it} = c_0 + Z(x_{it})' \beta + \delta_k(x_{it}' \theta_0) + \gamma_i + e_{it}.$$

Moreover, it is easy to check that the cardinality of P_m in Definition 2.1 is $\binom{m+d-1}{d-1}$, so the length of the vector $Z(x_{it})$ is $K = \frac{(d+k-1)!}{d!(k-1)!} - 1 = O(k^d)$. Thus, we may further write $Z(x_{it})$ as $Z(x_{it}) = (z_1(x_{it}), \dots, z_K(x_{it}))'$.

To eliminate the fixed effects, we take the within transformation and express the model in matrix form as

$$(I_N \otimes M_{i_T}) \mathcal{Y} = (I_N \otimes M_{i_T}) \mathcal{H}(\theta_0) \mathcal{C} + (I_N \otimes M_{i_T}) \mathcal{D}(\theta_0) + (I_N \otimes M_{i_T}) \mathcal{E} \quad (2.9)$$

$$= (I_N \otimes M_{i_T}) \mathcal{Z} \beta + (I_N \otimes M_{i_T}) \mathcal{D}(\theta_0) + (I_N \otimes M_{i_T}) \mathcal{E}, \quad (2.10)$$

where

$$\begin{aligned} \mathcal{Y}_{NT \times 1} &= (y_{11}, \dots, y_{1T}, \dots, y_{N1}, \dots, y_{NT})', \\ \mathcal{H}(\theta)_{NT \times (k-1)} &= (H(x'_{11}\theta), \dots, H(x'_{1T}\theta), \dots, H(x'_{N1}\theta), \dots, H(x'_{NT}\theta))' \text{ for } \forall \theta, \\ H(w)_{(k-1) \times 1} &= (h_1(w), h_2(w), \dots, h_{k-1}(w))' \text{ for } \forall w \in \mathbb{R}, \\ \mathcal{D}(\theta)_{NT \times 1} &= (\delta_k(x'_{11}\theta), \dots, \delta_k(x'_{1T}\theta), \dots, \delta_k(x'_{N1}\theta), \dots, \delta_k(x'_{NT}\theta))' \text{ for } \forall \theta, \\ \mathcal{Z}_{NT \times K} &= (Z(x_{11}), \dots, Z(x_{1T}), \dots, Z(x_{N1}), \dots, Z(x_{NT}))', \\ \mathcal{E}_{NT \times 1} &= (e_{11}, \dots, e_{1T}, \dots, e_{N1}, \dots, e_{NT})' \quad \text{and} \quad \mathcal{C}_{(k-1) \times 1} = (c_1, c_2, \dots, c_{k-1})'. \end{aligned} \quad (2.11)$$

Notice that c_0 , the location of $g(\cdot)$ on vertical direction, is also removed by the within transformation and thereby cannot be estimated by (2.9) and (2.10). To estimate c_0 , extra information is needed (e.g. Assumption 1.ix in Su and Jin (2012)). In Section 3 below, we shall propose a closed-form estimator $\hat{\theta}$ for θ_0 by (2.10). After that, we will plug $\hat{\theta}$ back to (2.9) and recover the link function in Section 4.

3 Closed-form estimator of the parameter θ_0

We consider in what follows a within ordinary least squares (OLS) estimator of β from (2.10):

$$\hat{\beta} = [\mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{Z}]^{-1} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{Y}. \quad (3.1)$$

To simplify the notation, for $i = 1, \dots, N$, let $Q_{1,i} = E[Z(x_{it})Z(x_{it})']$ and $q_i = E[Z(x_{it})]$. Denote further that $Q_1 = \frac{1}{N} \sum_{i=1}^N Q_{1,i}$, $\bar{q} = \frac{1}{N} \sum_{i=1}^N q_i$ and $Q_2 = \frac{1}{N} \sum_{i=1}^N q_i q_i'$. Moreover, for $t = 1, \dots, T$, let $x_t = (x_{1t}, \dots, x_{Nt})'$ and $e_t = (e_{1t}, \dots, e_{Nt})'$.

To proceed further, we introduce the following assumptions. A detailed discussion about the assumptions is given in the Appendix and their verifications are given in the supplementary document of this paper.

Assumption 1 (Covariates and errors):

- i. Let $E[e_{it}] = 0$ for all $1 \leq i \leq N$ and $1 \leq t \leq T$. Suppose that $\{(x_t, e_t) : 1 \leq t \leq T\}$ is strictly stationary and α -mixing. Denote $\alpha_{ij}(|t-s|)$ as the α -mixing coefficient between $\{x_{it}, e_{it}\}$ and $\{x_{js}, e_{js}\}$. For these coefficients, let $\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^{\infty} (\alpha_{ij}(t))^{\eta/(4+\eta)} = O(N)$ and $\sum_{i=1}^N \sum_{j=1}^N (\alpha_{ij}(0))^{\eta/(4+\eta)} = O(N)$, where $\eta > 0$ is chosen such that $E[|e_{it}|^{4+\eta}] < \infty$ and $E[\|x_{it}\|^{4+\eta}] < \infty$.
- ii. Suppose that x_{it} is independent of e_{js} for all $1 \leq i, j \leq N$ and $1 \leq t, s \leq T$.

Assumption 2 (Identifications):

- i. Let $\|\theta_0\| = 1$ and $\theta_{01} > 0$ where θ_{01} is the first element of θ_0 .
- ii. $E[g(x'_{it}\theta_0)] = 0$ for all $1 \leq i \leq N$ and $1 \leq t \leq T$. Moreover, for the same η in Assumption 1, let $E[|g(x'_{it}\theta_0)|^{2+\eta/2}] < \infty$.

Assumption 3 (Boundaries):

- i. Let the smallest eigenvalue of the $K \times K$ matrix $(Q_1 - Q_2)$ be uniformly bounded away from zero, such that $\lambda_{\min}(Q_1 - Q_2) > 0$ uniformly over K .
- ii. (1) There exists $r > 0$ such that $\sup_{0 \leq \epsilon \leq 1} \sup_{\|\theta - \theta_0\| \leq \epsilon} \max_{1 \leq i \leq N} E[\delta_k^2(x'_{it}\theta)] = O(k^{-r})$.
 (2) $\max_{1 \leq i \leq N} E[|\mathcal{H}_p(x_{i1})|^{4+\eta}] = O(|p|^d)$ as $|p| \rightarrow \infty$, where $|p|$ is given under (2.5).
- iii. $\frac{k^{3d}}{NT} \rightarrow 0$ and $\frac{k^{3d/2}}{T} \rightarrow 0$ as N, T, k tend to infinity jointly.

Under these conditions, we are ready to establish the main results of the section in Lemma 3.1 and Theorems 3.1 and 3.2; their proofs are given in the Appendix.

Lemma 3.1 *Let Assumptions 1, 2.i and 3 hold. As $(N, T) \rightarrow (\infty, \infty)$ jointly, $\|\hat{\beta} - \beta\|^2 = O_P\left(\frac{k^{3d/2}}{NT}\right) + O(k^{-r})$, where $r > 0$ is involved in Assumption 3.ii.(1).*

The current rate of convergence is similar to the second result of Theorem 1 in Newey (1997). Noting that $K = O(k^d)$, the first term of the rate is not $O_P\left(\frac{K}{NT}\right)$, which is due to the fact that we cannot bound the basis functions uniformly. To solve this type of problem, Newey (1997) puts restrictions on the basis functions. By contrast, we bound the moments of the basis functions, which is similar to Assumption 3.iv of Su and Jin (2012) but less restrictive. Moreover, the rate $O_P\left(\frac{K}{NT}\right)$ is achievable for the first term of the rate of convergence when the $(4 + \eta)^{\text{th}}$ order moments of the basis functions are bounded uniformly. This may be done in the same way as Assumption 3.iv of Su and Jin (2012). In practice, the uniform bound on the $(4 + \eta)^{\text{th}}$ order moments of the basis functions might be achieved by rescaling the data properly. In this case, one may have a faster rate of the order $O_P\left(\frac{K}{NT}\right)$ for the first term of the rate in Lemma 3.1. In the sequel, we shall establish the main results for the case of $c_1 \neq 0$ and then $c_1 = 0$, respectively.

The case of $c_1 \neq 0$. In view of (2.7), the first d elements of $\hat{\beta}$ are the estimates of the coefficients $c_1\theta_{01}, \dots, c_1\theta_{0d}$. Moreover, because of Lemma 3.1 and $\|\hat{\beta}^d - \beta^d\| \leq \|\hat{\beta} - \beta\|$, where $\hat{\beta}^d$ and β^d denote vectors of the first d elements of $\hat{\beta}$ and β , respectively, the estimates are consistent. Therefore, it follows from the identification condition in Assumption 2.i and the continuous mapping theorem that $\left(\sum_{i=1}^d \hat{\beta}_i^2\right)^{1/2} \rightarrow_P |c_1|$, where $\hat{\beta}_i$ denotes the i^{th} element of $\hat{\beta}$. Since $\theta_{01} > 0$ by Assumption 2.i again, $\hat{\beta}_1 \neq 0$ and $\hat{\beta}_1$ has the same sign as c_1 with probability approaching one, so that $\hat{c}_1 = \text{sgn}(\hat{\beta}_1) \left(\sum_{i=1}^d \hat{\beta}_i^2\right)^{1/2}$ is a consistent estimator of c_1 . Then, we define the estimator of θ_0 as

$$\hat{\theta} = \frac{1}{\hat{c}_1} Q_3 \hat{\beta}, \quad (3.2)$$

where $Q_3 = (I_d, 0_{d \times (K-d)})$ and hence $Q_3 \hat{\beta}$ chooses the first d elements of $\hat{\beta}$ automatically. The following theorem establishes an asymptotic normality for $\hat{\theta}$ of (3.2).

Theorem 3.1 *Let Assumptions 1, 2.i and 3 hold and $c_1 \neq 0$. Then $\hat{\theta}$ of (3.2) is consistent as $(N, T) \rightarrow (\infty, \infty)$ jointly. In addition, let $E \left[\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N Q_3 (Q_1 - Q_2)^{-1} (Z(x_{i1}) - q_i) e_{i1} \right\|^4 \right] = O(1)$, $\frac{NT}{k^r} \rightarrow 0$ and $\frac{k^{4.5d}}{NT} \rightarrow 0$. As $(N, T) \rightarrow (\infty, \infty)$ jointly, $\sqrt{NT} (\hat{\theta} - \theta_0) \rightarrow_D N(0, c_1^{-2} \cdot \Xi_0)$, where*

$$\Xi_0 = \lim_{N, T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T Q_3 (Q_1 - Q_2)^{-} \Sigma_{itjs} (Q'_1 - Q'_2)^{-} Q'_3,$$

with $\Sigma_{itjs} = E [e_{it} e_{js} (Z(x_{it}) - q_i) (Z(x_{js}) - q_j)']$.

The extra conditions required in the body of this theorem imply that achieving the asymptotic normality comes with a price such that $r > 4.5d$, which is caused by the second decomposition in (2.5) and can be considered as a trade-off in order to achieve the closed-form estimator. In the second part of the supplementary document of this paper, we provide some

detailed derivations to explain why the conditions: $\frac{NT}{k^r} \rightarrow 0$ and $\frac{k^{4.5d}}{NT} \rightarrow 0$ are reasonable. The restriction $E \left[\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N Q_3 (Q_1 - Q_2)^{-1} (Z(x_{i1}) - q_i) e_{i1} \right\|^4 \right] = O(1)$ is in the same spirit as Assumption ii of Lemma A.1 in Chen et al. (2012b) and is easily verified for the i.i.d case.

The case of $c_1 = 0$. Without loss of generality, there is some integer $m_0 > 1$ such that suppose that $c_{m_0} \neq 0$. The existence of m_0 only requires that the link function $g(\cdot)$ is not a constant function.

Noting by (2.8) that the coefficients for $h_{m_0}(x_{it,1}), \dots, h_{m_0}(x_{it,d})$ are $c_{m_0}\theta_{01}^{m_0}, \dots, c_{m_0}\theta_{0d}^{m_0}$, where θ_{0j} denotes the j^{th} element of θ_0 for $j = 1, \dots, d$. The corresponding estimates of these coefficients are the $m_1^{\text{th}}, \dots, m_d^{\text{th}}$ elements of $\hat{\beta}$ respectively, where $m_1 = \frac{(d+m_0-1)!}{d!(m_0-1)!}$, $m_2 = m_1 + \frac{(d+m_0-1)!}{(d-1)!m_0!} - \frac{(d+m_0-2)!}{(d-2)!m_0!}$, $m_3 = m_2 + \frac{(d+m_0-2)!}{(d-2)!m_0!} - \frac{(d+m_0-3)!}{(d-3)!m_0!}$, \dots , $m_d = \frac{(d+m_0)!}{d!m_0!} - 1$. Since m_0 is fixed, it follows from Lemma 3.1 and the continuous mapping theorem that $\left(\sum_{i=1}^d (\hat{\beta}_{m_i}^2)^{1/m_0} \right)^{m_0/2} \rightarrow_P |c_{m_0}|$ by Assumption 2.i. Hence, by the same reason as the case for $c_1 \neq 0$, a consistent estimator of c_{m_0} is $\hat{c}_{m_0} = \text{sgn}(\hat{\beta}_{m_1}) \left(\sum_{i=1}^d (\hat{\beta}_{m_i}^2)^{1/m_0} \right)^{m_0/2}$.

Meanwhile, because again $\hat{\beta}_{m_1}$ is the consistent estimate of $c_{m_0}\theta_{01}^{m_0}$, it follows from the continuous mapping theorem that $\hat{\theta}_{01} = (\hat{\beta}_{m_1}/\hat{c}_{m_0})^{1/m_0}$ is a consistent estimate for θ_{01} . Let us construct Q_3 in the same fashion as before such that $Q_3\hat{\beta}$ consists of the $m_1^{\text{th}}, (m_1+1)^{\text{th}}, \dots, (m_1+d-1)^{\text{th}}$ elements in $\hat{\beta}$. These elements are the estimates of $c_{m_0}\theta_{01}^{m_0}, \sqrt{m_0}c_{m_0}\theta_{01}^{m_0-1}\theta_{02}, \dots, \sqrt{m_0}c_{m_0}\theta_{01}^{m_0-1}\theta_{0d}$, respectively. With \hat{c}_{m_0} and $\hat{\theta}_{01}$ at hand, we are finally able to define an estimator for θ_0 of the form:

$$\hat{\theta} = Q_4 Q_3 \hat{\beta}, \quad \text{where } Q_4 = \text{diag} \left(\frac{1}{\hat{c}_{m_0} \hat{\theta}_{01}^{m_0-1}}, \frac{\hat{\theta}_{01}}{\sqrt{m_0} \hat{\beta}_{m_1}}, \dots, \frac{\hat{\theta}_{01}}{\sqrt{m_0} \hat{\beta}_{m_1}} \right). \quad (3.3)$$

The following theorem establishes another asymptotic normality for $\hat{\theta}$ of (3.3).

Theorem 3.2 *Suppose that c_{m_0} is the first nonzero coefficient given in (2.3) after $c_1 = 0$. Let Assumptions 1, 2.i and 3 hold. Then $\hat{\theta}$ of (3.3) is consistent as $(N, T) \rightarrow (\infty, \infty)$ jointly. In addition, let $\frac{NT}{k^r} \rightarrow 0$, $\frac{k^{4.5d}}{NT} \rightarrow 0$ and $E \left[\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N Q_3 (Q_1 - Q_2)^{-1} (Z(x_{i1}) - q_i) e_{i1} \right\|^4 \right] = O(1)$. As $(N, T) \rightarrow (\infty, \infty)$ jointly, $\sqrt{NT}(\hat{\theta} - \theta_0) \rightarrow_D N(0, Q\Xi_0 Q)$, where Ξ_0 is the same as in Theorem 3.1 and $Q = c_{m_0}^{-1} \theta_{01}^{-m_0+1} \text{diag}(1, m_0^{-1/2}, \dots, m_0^{-1/2})$.*

In the next section, we will investigate the estimation of the link function. In view of the above discussion on whether $c_1 = 0$, we will assume that $c_1 \neq 0$ and use $\hat{\theta}$ of (3.2) in the rest development of this paper due to similarity.

4 Estimation of the link function

As mentioned earlier, the constant c_0 cannot be estimated from the regression equations. Without c_0 the curve of the link function would be lowered c_0 unit in the vertical direction. The

stipulation of Assumption 2.ii is for the estimation of c_0 .

Lemma 4.1 Denote $\hat{c}_0 = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it})' \hat{\beta}$. Suppose that Assumptions 1–3 hold and let $r \geq 3d/2$. As $(N, T) \rightarrow (\infty, \infty)$ jointly, $(\hat{c}_0 - c_0)^2 = O_P\left(\frac{k^{3d}}{NT}\right) + O_P(k^{-r+3d/2})$.

In connection with (2.5) and Lemma 3.1, define an estimator for $g(x'\theta_0)$ for any $x \in \mathbb{R}^d$:

$$\hat{g}(x'\hat{\theta}) = Z(x)' \hat{\beta} + \hat{c}_0. \quad (4.1)$$

In view of the proof of Theorem 3.1, it is straightforward to have the following theorem and its proof is given in the supplementary document of this paper.

Theorem 4.1 Let Assumptions 1, 2 and 3 hold. In addition, let $\frac{NT}{k^r} \rightarrow 0$, $\frac{k^{4.5d}}{NT} \rightarrow 0$ and $E \left[\left\| \frac{1}{\sqrt{NK^{3/2}}} \sum_{i=1}^N (Z(x) - \bar{q})' (Q_1 - Q_2)^- (Z(x_{i1}) - q_i) e_{i1} \right\|^4 \right] = O(1)$. As $(N, T) \rightarrow (\infty, \infty)$ jointly, $\sqrt{\frac{NT}{K^{3/2}}} (\hat{g}(x'\hat{\theta}) - g(x'\theta_0)) \rightarrow_D N(0, \Xi_1)$, where $K = \frac{(d+k-1)!}{d!(k-1)!} - 1 = O(k^d)$ and

$$\Xi_1 = \lim_{N, T, k \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \frac{1}{K^{3/2}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \Sigma_{itjs} (Q'_1 - Q'_2)^- (Z(x) - \bar{q}),$$

with $\Sigma_{itjs} = E[e_{it} e_{js} (Z(x_{it}) - q_i) (Z(x_{js}) - q_j)']$.

Notice that the above theorem just recovers $g(x'\theta_0)$ rather than $g(w)$ itself. To estimate the link function $g(w)$ for any $w \in \mathbb{R}$, $\hat{\theta}$ is plugged in (2.9) and then the estimator of \mathcal{C} is obtained by OLS,

$$\hat{\mathcal{C}} = \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^- \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{Y}. \quad (4.2)$$

To simplify the notation, for each time series $\{x_{i1}, \dots, x_{iT}\}$ with $1 \leq i \leq N$, let $R_{1,i}(\theta) = E[H(x'_{it}\theta) H(x'_{it}\theta)']$ and $r_i(\theta) = E[H(x'_{it}\theta)]$. Denote that $R_1(\theta) = \frac{1}{N} \sum_{i=1}^N R_{1,i}(\theta)$, $\bar{r}(\theta) = \frac{1}{N} \sum_{i=1}^N r_i(\theta)$ and $R_2(\theta) = \frac{1}{N} \sum_{i=1}^N r_i(\theta) r_i(\theta)'$. Moreover, the next assumption is necessary for achieving the consistency and once again we leave the discussion to the Appendix.

Assumption 4:

- i. Let the smallest eigenvalue of the $(k-1) \times (k-1)$ matrix $(R_1(\theta) - R_2(\theta))$ be bounded away from zero uniformly on a neighborhood of θ_0 .
- ii. $\sup_{0 \leq \epsilon \leq 1} \sup_{\|\theta - \theta_0\| \leq \epsilon} \max_{1 \leq i \leq N} E[|h_m(x'_{i1}\theta)|^{4+\eta}] = O(m)$ as $m \rightarrow \infty$, where η is given in Assumption 1.
- iii. Suppose that x_{it} has a support $\mathbb{X} \subset \mathbb{R}^d$. For $\forall x \in \mathbb{X}$, $g(x'\theta)$ satisfies a Lipschitz condition on a neighborhood of θ_0 , U_{θ_0} , such that $|g(x'\theta_1) - g(x'\theta_0)| \leq M(x) \|\theta_1 - \theta_0\|$, where $\theta_1 \in U_{\theta_0}$. Moreover, $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (M(x_{it}))^2 = O_P(1)$.

Now, we are able to establish the following lemma.

Lemma 4.2 *Under the conditions of Theorem 3.1 and Assumption 4, as $(N, T) \rightarrow (\infty, \infty)$ jointly, $\|\hat{\mathcal{C}} - \mathcal{C}\|^2 = O_P\left(\frac{k^{3/2}}{NT}\right) + O_P(k^{-r})$.*

Similar to the discussion under Lemma 3.1, if we use a stronger assumption to bound the $(4 + \eta)^{\text{th}}$ order moment of $h_m(x'_{it}\theta)$ uniformly, then the first term of the rate of convergence above would become $O_P\left(\frac{k}{NT}\right)$. Note also that alternatively the constant c_0 may be estimated invoking $\hat{\mathcal{C}}$ instead of $\hat{\beta}$: $\tilde{c}_0 = -\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T H\left(x'_{it}\hat{\theta}\right)' \hat{\mathcal{C}}$. Moreover, by a similar procedure to the proof of Lemma 4.1, $|\tilde{c}_0 - c_0|^2 = O_P\left(\frac{k^3}{NT}\right)$. Consequently, we define an estimator for $g(w)$:

$$\tilde{g}(w) = H(w)' \hat{\mathcal{C}} + \tilde{c}_0, \quad (4.3)$$

where $H(w)$ is denoted in (2.11). Then the accuracy of the nonparametric estimate is given below.

Theorem 4.2 *Let Assumption 2.ii hold. Under the conditions of Lemma 4.2, as $(N, T) \rightarrow (\infty, \infty)$ jointly, $\|\tilde{g}(w) - g(w)\|_{L^2}^2 = O_P\left(\frac{k^3}{NT}\right) + O\left(\|\delta_k(w)\|_{L^2}^2\right)$ where for $f(w) \in L^2\left(\mathbb{R}, e^{-w^2/2}\right)$, $\|f(w)\|_{L^2}^2 = \int_{\mathbb{R}} f^2(w) \exp(-w^2/2) dw$ is the metric in the Hilbert space.*

Note that we may have an optimal choice for k when $\|\delta_k(w)\|_{L^2}^2$ has the same order as $\frac{k^3}{NT}$. This is similar to the situation in the literature (see, for example, Condition 3.5.ii in Chen (2007) and Assumption 2 of Newey and Powell (2003)). More details and discussion are provided in the supplementary document.

We will evaluate the proposed model and the estimation method using both simulated and real data examples in Section 5 below.

5 Numerical study

In this section, we provide a Monte Carlo simulation and one empirical study. In the simulation, we consider an exponential function $g(w) = \exp(w)$ to be a link function. Certainly, this violates Assumption 2.ii that $E[g(x'_{it}\theta_0)] = 0$ stipulated for the estimate of c_0 , but it does not affect the estimation of θ_0 . A similar discussion can be found in the Monte Carlo study section of Su and Jin (2012). Meanwhile, a simple calculation yields $c_1 = \sqrt{e} \neq 0$ so that $\hat{\theta}$ is constructed by the first d elements of $\hat{\beta}$. In the empirical study, we investigate the demand for cigarettes in the US.

5.1 Monte Carlo experiments

The data generating process (DGP) is as follows.

$$y_{it} = \exp(x_{1,it}\theta_{01} + x_{2,it}\theta_{02}) + \gamma_i + e_{it} \quad (5.1)$$

and for $j = 1$ and 2 and $t = -99, \dots, 0, \dots, T$, $x_{j,it} = \rho_{x_j} x_{j,i,t-1} + i.i.d. N(0, 1)$ where $\rho_{x_1} = 0.7$, $\rho_{x_2} = 0.3$, $x_{j,i,-100} = 0$. To introduce a type of cross-sectional dependence to the model, we follow the DGP in Chen et al. (2012a) by letting $e_t = (e_{1t}, \dots, e_{Nt})'$, $e_{-100} = 0_{N \times 1}$ and $\rho_e = 0.2$ for $1 \leq t \leq T$. Then the error term e_t is generated as $e_t = \rho_e e_{t-1} + i.i.d. N(0_{N \times 1}, \Sigma_e)$ for $t = -99, \dots, 0, \dots, T$, where $\Sigma_e = (\sigma_{ij})_{N \times N}$ with $\sigma_{ij} = 0.5^{|i-j|}$ for $1 \leq i, j \leq N$.

To see whether the dependence of the fixed effects on the regressor affects our proposed method, we consider two situations of the generation of γ_i : (1) $\gamma_i \sim i.i.d. U(0, 1)$ and is independent of the regressors; (2) $\gamma_i = x_{1,i1} + x_{2,i1}$ for all i .

Set $\theta_{01} = 0.8$ and $\theta_{02} = -0.6$. To meet the theoretical requirement we take the truncation parameter $k = \lfloor aN^{\rho_1} T^{\rho_2} \rfloor$ with $a = 1$, $\rho_1 = \rho_2 = 1/5$ to ensure that k changes with N and T ; we also do the simulations for fixed $k = 5$ but report the results in the supplement. See Su and Jin (2012) and Dong et al. (2014) for a similar treatment. As a comparison, we use the profile estimate method (see, for example, Chen (2007) and Dong et al. (2014)) as well. To start with the profile method, we use the true value θ_0 . We also tried to use the within-OLS estimate on the linear model (5.3) below as an initial value, but the results of the profile method do not show consistency. We repeat the estimation procedure 1000 times. For each replication, we record the bias and squared error as: $bias = \hat{\theta}_j - \theta_{j0}$ and $se = (\hat{\theta}_j - \theta_{j0})^2$ for $j = 1, 2$, where $\hat{\theta}_j$ denotes the estimate of θ_{0j} . After 1000 replications, we report the mean of these biases and the root of the mean of these squared errors, which are labeled as Bias and RMSE in Table 1.

As can be seen from Tables 1 and 2, all the results are very accurate, in the sense that the Biases and RMSEs of $\hat{\theta}_1$ and $\hat{\theta}_2$ decrease as both N and T increase. It seems that the rates of the RMSEs of the closed-form estimate decrease to 0 are faster than those from the profile method in most cases and that the fixed effect dependence has less impact on our method than the profile method. Note that the bias of the closed-form estimate always performs extremely well. This is possibly due to the high-order differentiability of the link function and the choice of the distribution of x_{it} .

5.2 Empirical study: US cigarette demand

In this section, we provide an empirical study to show how our method is implemented in practice by considering the US cigarette demand.

The data set of this case study is from Baltagi et al. (2000) for analyzing the demand for cigarettes in the US. C_{it} is the real per capita sales of cigarettes (measured in packs), DI_{it} is the real per capita disposable income, P_{it} is the average retail price of a pack of cigarettes measured in real terms, PN_{it} is the minimum real price of cigarettes in any neighboring state and u_{it} is the disturbance term. $i = 1, \dots, 46$ and $t = 1, \dots, 30$ represent the states and the years (1963–1992) respectively.

			Closed-form estimator			Profile method		
	$k = \lfloor (NT)^{1/5} \rfloor$	$N \setminus T$	30	60	100	30	60	100
$\hat{\theta}_1$	Bias	30	-0.0024	-0.0009	-0.0006	-0.0120	-0.0120	-0.0105
		60	-0.0017	0.0004	0.0001	-0.0104	-0.0093	-0.0089
		100	-0.0003	0.0001	-0.0001	-0.0115	-0.0091	-0.0052
	RMSE	30	0.0210	0.0181	0.0161	0.0227	0.0182	0.0163
		60	0.0194	0.0086	0.0067	0.0170	0.0130	0.0120
		100	0.0175	0.0072	0.0039	0.0174	0.0125	0.0062
$\hat{\theta}_2$	Bias	30	-0.0022	0.0000	-0.0002	-0.0013	-0.0013	0.0001
		60	0.0020	0.0007	0.0002	0.0007	-0.0001	-0.0002
		100	0.0004	0.0002	-0.0001	-0.0010	-0.0001	-0.0001
	RMSE	30	0.0276	0.0233	0.0210	0.0397	0.0258	0.0229
		60	0.0251	0.0115	0.0090	0.0261	0.0149	0.0140
		100	0.0241	0.0095	0.0053	0.0243	0.0140	0.0059

Table 1: Bias and RMSE for the case where $\gamma_i \sim i.i.d.U(0, 1)$

		Closed-form estimator				Profile method		
	$k = \lfloor (NT)^{1/5} \rfloor$	$N \setminus T$	15	30	80	15	30	80
$\hat{\theta}_1$	Bias	20	-0.0043	-0.0032	-0.0014	-0.0135	-0.0130	-0.0115
		50	-0.0020	-0.0022	-0.0011	-0.0121	-0.0118	-0.0111
		100	0.0001	-0.0002	0.0002	-0.0103	-0.0110	-0.0102
	RMSE	20	0.0369	0.0251	0.0160	0.0305	0.0262	0.0209
		50	0.0237	0.0160	0.0104	0.0235	0.0192	0.0159
		100	0.0175	0.0120	0.0078	0.0189	0.0174	0.0144
$\hat{\theta}_2$	Bias	20	-0.0027	-0.0028	-0.0013	-0.0026	-0.0017	0.0001
		50	-0.0013	-0.0023	-0.0012	-0.0003	-0.0013	-0.0010
		100	0.0009	0.0000	0.0004	0.0021	0.0008	0.0011
	RMSE	20	0.0474	0.0328	0.0214	0.0541	0.0458	0.0337
		50	0.0324	0.0212	0.0138	0.0439	0.0306	0.0211
		100	0.0235	0.0161	0.0104	0.0334	0.0253	0.0175

Table 2: Bias and RMSE for the case where $\gamma_i = x_{1,i1} + x_{2,i1}$

We consider fitting the data by a semiparametric single-index model of the form

$$y_{it} = g(x'_{it}\theta_0) + \gamma_i + e_{it}, \quad (5.2)$$

where $y_{it} = \ln C_{it}$, $x_{it} = (\ln C_{i,t-1}, \ln DI_{it}, \ln P_{it}, \ln PN_{it})'$. Due to the lagged dependent value included in x_{it} , the length of time series used in the regression is 29 (such that $t = 2, \dots, 30$). $\{\gamma_i\}$ captures all the state-specific effects. All the cross-sectional dependences of the errors and year-specific effects are reflected in e_{it} . As a comparison, we also run the within-OLS

regression on the linear model given below:

$$y_{it} = x'_{it}\theta_0 + \gamma_i + e_{it}. \quad (5.3)$$

In the following, SIM and LIM denote the single-index model (5.2) and the linear model (5.3), respectively.

To use the proposed method for model (5.2), the choice of an optimal truncation parameter k is an open question. To choose k , we thus use an extended version of the generalized cross-validation (GCV) criterion proposed in Gao et al. (2002). Specifically, we calculate the GCV for each time series by (3.3) of Gao et al. (2002) and then take an average for these GCVs. These GCVs suggest a data-driven \hat{k} that corresponds to the minimum GCV. We then consider some possible k values around the \hat{k} and calculate the mean square error (MSE) for the corresponding models. The in-sample mean squared error (In-MSE) and out-sample mean squared error (Out-MSE) are employed to measure the performance of the SIM with different k values and the LIM.

To get In-MSE, all the data collected above are used to run the regression in order to get $\hat{\beta}_{\text{In}}$. Then In-MSE is $\text{In-MSE} = \frac{1}{46 \times 29} \sum_{i=1}^{46} \sum_{t=1}^{29} (\tilde{y}_{it} - \tilde{Z}(x_{it})' \hat{\beta}_{\text{In}})^2$, where $\tilde{y}_{it} = y_{it} - \frac{1}{29} \sum_{t=1}^{29} y_{it}$ and $\tilde{Z}(x_{it})$ is defined in the same fashion. To get Out-MSE, only part of data ($i = 1, \dots, 46$, $t = 1, \dots, \tilde{T}$ and $\tilde{T} = 25, \dots, 28$) are used to estimate $\hat{\beta}_{\text{Out}, \tilde{T}}$ in order to forecast $\tilde{y}_{i, \tilde{T}+1}$. Then Out-MSE is obtained as $\text{Out-MSE} = \frac{1}{46 \times (29 - 25)} \sum_{i=1}^{46} \sum_{\tilde{T}=25}^{28} (\tilde{y}_{i, \tilde{T}+1} - \tilde{Z}(x_{i, \tilde{T}+1})' \hat{\beta}_{\text{Out}, \tilde{T}})^2$, where $\tilde{y}_{i, \tilde{T}+1} = y_{i, \tilde{T}+1} - \frac{1}{\tilde{T}+1} \sum_{t=1}^{\tilde{T}+1} y_{it}$ and $\tilde{Z}(x_{i, \tilde{T}+1})$ is defined in the same fashion.

	SIM			LIM
	$k = 2$	$k = 3$	$k = 4$	
In-MSE	0.00137	0.00123	0.00117	0.00137
Out-MSE	0.00180	0.00176	0.00209	0.00180
GCV	0.00158	0.00152	0.00157	

Table 3: GCV, In-MSE and Out-MSE

SIM	\hat{k}	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$
	3	0.151	0.678	0.212	-0.687
		(0.214)	(0.148)	(0.318)	(0.303)

Table 4: Estimated coefficients for US cigarette demand

Note that as $k = 2$ (this means that the Hermite polynomials h_0 and h_1 are chosen) the single-index model will reduce to the linear model. This is why the In-MSE and Out-MSE for $k = 2$ are the same as the results of the linear model in Table 3. As Table 3 shows, both GCV and Out-MSE indicate that the optimal $k = 3$, while the In-MSE has a weak favor on $k = 4$. Thus, we choose $k = 3$ for the rest calculation (such that functions h_0 , h_1 and h_2 are

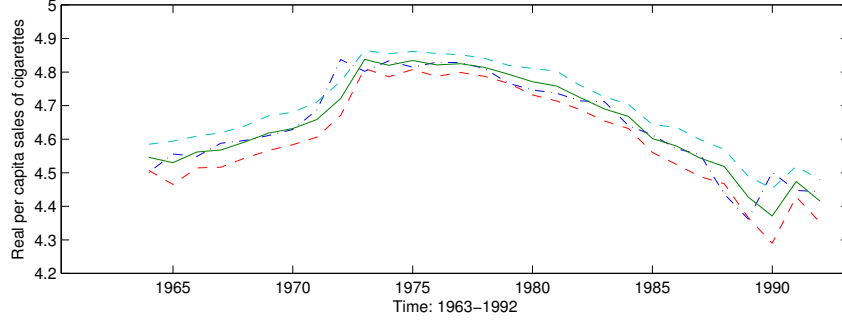


Figure 1: Estimated log real per capita sales of cigarettes

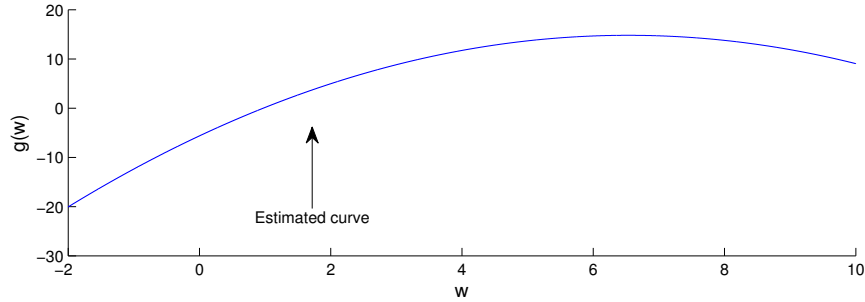


Figure 2: Estimated curve for US cigarette demand

chosen). The estimates of the coefficients are reported in Table 4. Due to the similarity, we only give the plots for one state in Figure 1 and omit the others. The dash-dot line is the real per capita sales of cigarettes, the solid line is the estimated per capita sales of cigarettes by our approach, and the two dash lines are the 95% confidence bands. Figure 2 clearly shows that the estimated curve appears to be nonlinear.

6 Conclusion and discussion

In this paper, we have considered a semiparametric single-index panel data model allowing for cross-sectional dependence, stationarity and unobservable heterogeneity. Some closed-form estimates have been proposed to recover the parameters of interest and the link function. The resulting asymptotic theory has been established and illustrated using both simulated and empirical examples. As both the theory and Monte Carlo study have suggested, our estimation method performs well when cross-sectional dependence exists in the system. Moreover, since we have not imposed any specific assumption on the fixed effects, they can be correlated with the regressors to capture unobservable heterogeneity. One empirical example has shown that the proposed estimation method outperforms some natural competitors.

Note that the issue of whether the proposed parameter estimator is semiparametrically efficient is an important one. Meanwhile, the issue of whether we may consider a generalized

partially linear single-index panel data setting of the model discussed in Carroll et al. (1997) is another important one. Note also that the asymptotic theory of the estimator of θ_0 relies on the knowledge that whether $c_1 \neq 0$ in the orthogonal series expansion. In view of the proof of Theorem 3.1, this is testable, because we may establish an asymptotic normality for $Q_3\hat{\beta}$, where the matrix Q_3 chooses the elements in $\hat{\beta}$ containing c_1 and the hypothesis $c_1 = 0$ is equivalent to that of the vector $Q_3\hat{\beta} = 0$ jointly. We wish to leave such issues for future research.

7 Acknowledgments

The authors would like to thank the Co-Editor, Professor Jianqing Fan, the Associate Editor and the referees for their constructive suggestions and comments on an earlier version of this paper. The authors also acknowledge the Australian Research Council Discovery Grants Program support under Grant numbers: DP1096374 and DP130104229.

Appendix A Discussion on the assumptions

Assumption 1.i requires stationarity for the time series dimension. We use the α -mixing coefficient to measure the relationship between $\{x_{it}, e_{it}\}$ and $\{x_{js}, e_{js}\}$. This set-up is in the same spirit as Assumption A2 of Chen et al. (2012a) and Assumption C of Bai (2009). Since the mixing properties are hereditary, it allows us to avoid imposing restrictions on the functions. Thus, possible cross-sectional dependence and serial dependence are both captured by the spatial mixing coefficients. Particularly, $\alpha_{ij}(0)$ only measures the cross-sectional dependence between $\{x_{it}, e_{it}\}$ and $\{x_{jt}, e_{jt}\}$. A detailed example is given in the supplementary document of this paper.

Assumption 1.ii can be further relaxed by using a condition similar to Assumption 3.2.ii of Gao and Phillips (2013). Since we focus on establishing closed-form estimates for the stationary panel data single-index model considered in this paper, stationarity is an essential assumption. When x_{it} is non-stationary, the endogeneity may come in the system by using the technique provided in Wang and Phillips (2009) and Dong and Gao (2014). This is beyond the scope of this paper. We will leave it for future study.

Assumption 2.i is a standard identification requirement. Alternatively, one may follow Ichimura (1993) to assume $\theta_{01} = 1$. Assumption 2.ii is not really necessary when the main interests are on estimating the parameter θ_0 and measuring the changes in output y . Assumption 2.ii kicks in only if the location of a curve needs to be estimated. In practice, the true expectation of $E[g(x'_{it}\theta_0)]$ may not be zero, so Assumption 2.ii essentially means that one may estimate $g(x'_{it}\theta_0) - E[g(x'_{it}\theta_0)]$ rather than $g(x'_{it}\theta_0)$ (c.f. Su and Jin (2012)). An example is given in the Monte Carlo study for illustration.

Assumption 3.i can be verified by carrying out a procedure similar to Lemma A.2 in Gao et al. (2002) and it is also similar to Assumption 2 in Newey (1997) and Assumption 3.iv in Su and Jin (2012). Assumption 3.ii is similar to Assumptions 2.ii and 3 in Newey (1997) for the cross-sectional setting. In the first condition of this assumption, r is highly related to the smoothness of the link function and the second condition is more general than Assumption 3.iv in Su and Jin (2012) in the

panel data setting. The relevant discussion and detailed examples on the first condition can be found in Lemma A.1 of Dong et al. (2014). By the same argument as used below Assumption 2 in Newey (1997), it is not possible to assume $\mathcal{H}_p(x_{i1})$ is bounded uniformly. Therefore, we put restrictions on the moments of the basis functions. Compared with putting the bounds on the basis power series directly like Newey (1997), we believe that our current assumption is more realistic. Detailed examples are given in the supplementary document of this paper to demonstrate our purpose on this assumption.

In practice, the lengths of the cross-sectional dimension and time series dimension can be relatively large, so Assumption 3.iii is easy to achieve. Moreover, the researcher normally assumes that $N/T \rightarrow c \in (0, \infty]$ as $(N, T) \rightarrow (\infty, \infty)$ in the conventional panel data case, which is also covered by Assumption 3.iii.

Assumptions 4.i–ii are in the same spirit as Assumptions 3.i–ii. Assumption 4.iii is similar to Assumptions 5.3.1 and 5.5 in Ichimura (1993) and Assumption 3 in Newey and Powell (2003) in the cross-sectional setting. In the literature of the sieve estimation, Assumption 4.iii is equivalent to the “envelop condition”. We impose the Lipschitz condition on a neighborhood of θ_0 rather than requiring \mathbb{X} to be compact. In this sense, this assumption is more general than what has been assumed in Ichimura (1993). The last equation in Assumption 4.iii can be easily verified under certain restrictions by following a procedure similar to the second result of Lemma B.2 in the Appendix.

Appendix B Proof of the main results

Firstly, we give three useful lemmas before we prove the main results. The proofs of the lemmas are provided in the supplementary document.

Lemma B.1 *Suppose that $u = (u_1, \dots, u_d)'$, $v = (v_1, \dots, v_d)' \in \mathbb{R}^d$ and $\|v\| = 1$. Then*

$$H_m(u'v) = \sum_{|p|=m} \binom{m}{p} \prod_{j=1}^d H_{p_j}(u_j) \prod_{j=1}^d v_j^{p_j},$$

where $p = (p_1, \dots, p_d)$, p_j for $j = 1, \dots, d$ are all non-negative integers, $|p| = p_1 + \dots + p_d$ and $\binom{m}{p} = \frac{m!}{\prod_{j=1}^d p_j!}$.

This is Lemma 12.4.2 of Blower (2009) but its proof in the book is not correct.

Lemma B.2 *Let Assumptions 1, 2 and 3 hold. Then, as (N, T) go to (∞, ∞) jointly:*

1. $E \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it}) Z(x_{it})' - Q_1 \right\|^2 = O\left(\frac{k^{3d}}{NT}\right);$
2. $E \left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T g(x'_{it} \theta_0) \right]^2 = O\left(\frac{1}{NT}\right);$
3. $E \left\| \frac{1}{T} \sum_{t=1}^T Z(x_{it}) - q_i \right\|^2 = O\left(\frac{k^{3d/2}}{T}\right);$
4. $E \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it}) e_{it} \right\|^2 = O\left(\frac{k^{3d/2}}{NT}\right);$
5. $\lambda_{\min} \left(\frac{1}{NT} Z' (I_N \otimes M_{iT}) Z \right) \geq \lambda_{\min} (Q_1 - Q_2) / 2 > 0.$

Lemma B.3 *Let Assumptions 1–4 hold. Then the following results hold uniformly in a small neighbor of θ_0 as (N, T) go to (∞, ∞) jointly:*

1. $E \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T H(x'_{it}\theta) H(x'_{it}\theta)' - R_1(\theta) \right\|^2 = O\left(\frac{k^3}{NT}\right);$
2. $E \left\| \frac{1}{T} \sum_{t=1}^T H(x'_{it}\theta) - r_i(\theta) \right\|^2 = O\left(\frac{k^{3/2}}{T}\right);$
3. $E \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T H(x'_{it}\theta) e_{it} \right\|^2 = O\left(\frac{k^{3/2}}{NT}\right);$
4. $\lambda_{\min} \left(\frac{1}{NT} \mathcal{H}(\theta)' (I_N \otimes M_{i_T}) \mathcal{H}(\theta) \right) \geq \lambda_{\min} (R_1(\theta) - R_2(\theta)) / 2 > 0.$

For the sake of the representation, we introduce the following notation:

$$\begin{aligned} \bar{y}_i &= \frac{1}{T} \sum_{t=1}^T y_{it}, \quad \bar{H}_i(\theta_0) = \frac{1}{T} \sum_{t=1}^T H(x'_{it}\theta_0), \quad \bar{Z}_i = \frac{1}{T} \sum_{t=1}^T Z(x_{it}), \\ \bar{\delta}_{k,i}(\theta_0) &= \frac{1}{T} \sum_{t=1}^T \delta_k(x'_{it}\theta_0), \quad \bar{e}_i = \frac{1}{T} \sum_{t=1}^T e_{it}, \\ \tilde{y}_{it} &= y_{it} - \bar{y}_i, \quad \tilde{H}_{it}(\theta_0) = H(x'_{it}\theta_0) - \bar{H}_i(\theta_0), \quad \tilde{Z}_{it} = Z(x_{it}) - \bar{Z}_i, \\ \tilde{\delta}_k(x'_{it}\theta_0) &= \delta_k(x'_{it}\theta_0) - \bar{\delta}_{k,i}(\theta_0), \quad \tilde{e}_{it} = e_{it} - \bar{e}_i. \end{aligned}$$

Proof of Lemma 3.1:

By the uniqueness of the Moore–Penrose inverse and the fifth result of Lemma B.2 of this Appendix, the $K \times K$ matrix $\left[\frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{Z} \right]^-$ is the inverse of $\frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{Z}$ for each K . Therefore,

$$\begin{aligned} \hat{\beta} - \beta &= \left[\mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{Z} \right]^- \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{E} \\ &\quad + \left[\mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{Z} \right]^- \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{D}(\theta_0). \end{aligned} \quad (\text{B.1})$$

Focusing on $\frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{E}$ firstly, we have

$$E \left\| \frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{E} \right\|^2 \leq 2E \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it}) e_{it} \right\|^2 + 2E \left\| \frac{1}{N} \sum_{i=1}^N \bar{Z}_i \bar{e}_i \right\|^2. \quad (\text{B.2})$$

By the fourth result of Lemma B.2, we have $E \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it}) e_{it} \right\|^2 = O\left(\frac{k^{3d/2}}{NT}\right)$. For the second term on RHS of (B.2), write

$$\begin{aligned} E \left\| \frac{1}{N} \sum_{i=1}^N \bar{Z}_i \bar{e}_i \right\|^2 &= \sum_{u=1}^K E \left[\frac{1}{NT^2} \sum_{i=1}^N \sum_{t_1=1}^T \sum_{t_2=1}^T z_u(x_{it_1}) e_{it_2} \right]^2 \\ &= \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N E \left[\frac{1}{T^2} \sum_{t_2=1}^T \sum_{t_4=1}^T e_{it_2} e_{jt_4} \right] \cdot \sum_{u=1}^K E \left[\frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_3=1}^T z_u(x_{it_1}) z_u(x_{jt_3}) \right], \end{aligned}$$

where the last line follows from Assumption 1.ii. By Cauchy–Schwarz inequality, moment monotonicity and Assumption 3.ii respectively, we have

$$\left| \sum_{u=1}^K E \left[\frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_3=1}^T z_u(x_{it_1}) z_u(x_{jt_3}) \right] \right| \leq \sum_{u=1}^K \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_3=1}^T (E[z_u^2(x_{it_1})] E[z_u^2(x_{jt_3})])^{1/2}$$

$$\begin{aligned}
&\leq \sum_{u=1}^K \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_3=1}^T (E[z_u^{4+\eta}(x_{it_1})] E[z_u^{4+\eta}(x_{jt_3})])^{1/(4+\eta)} \\
&\leq \sum_{u=1}^K \frac{1}{T^2} \sum_{t_1=1}^T \sum_{t_3=1}^T (E[z_u^{4+\eta}(x_{it_1})] E[z_u^{4+\eta}(x_{jt_3})])^{1/4} = O(k^{3d/2}).
\end{aligned}$$

Similar to the proof of the first result of Lemma B.2, $\frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t_2=1}^T \sum_{t_4=1}^T E[e_{it_2} e_{jt_4}] = O\left(\frac{1}{NT}\right)$. Thus, $E\left\|\frac{1}{N} \sum_{i=1}^N \bar{Z}_i \cdot \bar{e}_i\right\|^2 = O\left(\frac{k^{3d/2}}{NT}\right)$. Thus, we have shown

$$\left\|\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{E}\right\| = O_P\left(\frac{k^{3d/4}}{\sqrt{NT}}\right). \quad (\text{B.3})$$

According to the fifth result of Lemma B.2 and (B.3), we obtain

$$\begin{aligned}
&\left\|[\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{E}\right\|^2 \\
&= \mathcal{E}'(I_N \otimes M_{i_T}) \mathcal{Z} [\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}]^- [\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{E} \\
&\leq \left[\lambda_{\min}\left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}\right)\right]^{-2} \cdot \left\|\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{E}\right\|^2 = O_P\left(\frac{k^{3d/2}}{NT}\right). \quad (\text{B.4})
\end{aligned}$$

We now consider $[\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{D}(\theta_0)$ and write

$$\begin{aligned}
&E\left\|(\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z})^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{D}(\theta_0)\right\|^2 \\
&= E\left[\mathcal{D}(\theta_0)'(I_N \otimes M_{i_T}) \mathcal{Z} (\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}/(NT))^- \right. \\
&\quad \left. \cdot (\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z})^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{D}(\theta_0)/(NT)\right] \\
&\leq E\left[\left[\lambda_{\min}(\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}/(NT))\right]^{-1} \right. \\
&\quad \left. \cdot \mathcal{D}(\theta_0)'(I_N \otimes M_{i_T}) \mathcal{Z} (\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z})^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{D}(\theta_0)/(NT)\right] \\
&\leq E\left[\left[\lambda_{\min}(\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}/(NT))\right]^{-1} \cdot \lambda_{\max}(W) \cdot \|\mathcal{D}(\theta_0)\|^2/(NT)\right], \quad (\text{B.5})
\end{aligned}$$

where the first inequality follows from the exercise 5 on page 267 of Magnus and Neudecker (2007) (or (A.6) in Su and Jin (2012)). Note that $W = (I_N \otimes M_{i_T}) \mathcal{Z} (\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z})^- \mathcal{Z}'(I_N \otimes M_{i_T})$ is symmetric and idempotent, so $\lambda_{\max}(W) = 1$. By Assumption 3.ii, it is easy to know that $E[\|\mathcal{D}(\theta_0)\|^2/(NT)] = O(k^{-r})$. In connection with the fifth result of Lemma B.2 of this Appendix, we obtain that the norm of the second term on RHS of (B.1) is $O_P(k^{-r/2})$, which further implies

$$\left\|[\mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T}) \mathcal{D}(\theta_0)\right\|^2 = O_P(k^{-r}). \quad (\text{B.6})$$

Therefore, the result follows from (B.4) and (B.6). ■

Proof of Theorem 3.1:

The consistency is a consequence of Lemma 3.1. We only show the normality in the sequel. It follows from Lemma 3.1 and the continuous mapping theorem that $\left(\sum_{i=1}^d \hat{\beta}_i^2\right)^{1/2} \rightarrow_P |c_1|$ in probability,

so we need only to consider $\sqrt{NT}Q_3(\hat{\beta} - \beta)$ and write

$$\begin{aligned}\sqrt{NT}Q_3(\hat{\beta} - \beta) &= \sqrt{NT}Q_3[\mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{D}(\theta_0) \\ &\quad + \sqrt{NT}Q_3[\mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{E}.\end{aligned}\quad (\text{B.7})$$

Note that $K = O(k^d)$ and $Q_3 = O(1)$. In connection with (B.6) and the assumption in the body of this theorem, it is straightforward to obtain

$$\left\| \sqrt{NT} \cdot Q_3[\mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{D}(\theta_0) \right\| \leq \sqrt{NT} \cdot O_P(k^{-r/2}) = o_P(1). \quad (\text{B.8})$$

Then, to achieve the normality, we need only to consider the second term on RHS of (B.7):

$$\begin{aligned}& \sqrt{NT} \cdot Q_3[\mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z}]^- \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{E} \\ &= \sqrt{NT} \cdot Q_3 \left(\left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z} \right)^- - (Q_1 - Q_2)^- \right) \left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{E} \right) \\ &\quad + \sqrt{NT} \cdot Q_3(Q_1 - Q_2)^- \left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{E} \right)\end{aligned}\quad (\text{B.9})$$

For two non-singular symmetric matrices A, B with same dimensions, we observe that

$$\begin{aligned}\|A^{-1} - B^{-1}\|^2 &= \|B^{-1}(B - A)A^{-1}\|^2 = \|\text{vec}(B^{-1}(B - A)A^{-1})\|^2 \\ &= \|(A^{-1} \otimes B^{-1})\text{vec}(B - A)\|^2 \leq \lambda_{\min}^{-2}(A \otimes B) \|\text{vec}(B - A)\|^2 = \lambda_{\min}^{-2}(A \otimes B) \|B - A\|^2,\end{aligned}$$

where $\lambda_{\min}(A \otimes B) = \lambda_{\min}(A) \cdot \lambda_{\min}(B)$. Therefore, in connection with the proof of the fifth result of Lemma B.2, $\left\| \left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z} \right)^- - (Q_1 - Q_2)^- \right\| = O_P\left(\sqrt{\frac{k^{3d}}{NT}}\right)$. Moreover, by (B.3), we obtain

$$\begin{aligned}& \left\| \sqrt{NT} \cdot Q_3 \left(\left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{Z} \right)^- - (Q_1 - Q_2)^- \right) \left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{E} \right) \right\| \\ &\leq \sqrt{NT} \cdot O_P\left(\sqrt{\frac{k^{3d}}{NT}}\right) \cdot O_P\left(\sqrt{\frac{k^{3d/2}}{NT}}\right) = O_P\left(\sqrt{\frac{k^{4.5d}}{NT}}\right) = o_P(1).\end{aligned}$$

The second term on RHS of (B.9) can be written as follows:

$$\begin{aligned}& \sqrt{NT} \cdot Q_3(Q_1 - Q_2)^- \left(\frac{1}{NT} \mathcal{Z}'(I_N \otimes M_{i_T})\mathcal{E} \right) \\ &= \sqrt{NT} \cdot Q_3(Q_1 - Q_2)^- \left(\frac{1}{N} \sum_{i=1}^N (q_i - \bar{Z}_{i.}) \bar{e}_{i.} + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (Z(x_{it}) - q_i) e_{it} \right).\end{aligned}\quad (\text{B.10})$$

Note also that $E \left\| \frac{1}{N} \sum_{i=1}^N (q_i - \bar{Z}_{i.}) \bar{e}_{i.} \right\|^2 = O \left(\frac{k^{3d/2}}{NT^2} \right)$. Similar to (B.4), it is easy to show that $\left\| (Q_1 - Q_2)^- \frac{1}{N} \sum_{i=1}^N (q_i - \bar{Z}_{i.}) \bar{e}_{i.} \right\| = O_P \left(\sqrt{\frac{k^{3d/2}}{NT^2}} \right)$. Therefore, we have

$$\left\| \sqrt{NT} \cdot Q_3 (Q_1 - Q_2)^- \frac{1}{N} \sum_{i=1}^N (q_i - \bar{Z}_{i.}) \bar{e}_{i.} \right\| \leq \sqrt{NT} \cdot O(1) \cdot O_P \left(\sqrt{\frac{k^{3d/2}}{NT^2}} \right) = o_P(1).$$

Since x_{it} and e_{it} are assumed to be stationary and α -mixing, we now use the large-block and small-block technique (e.g. Theorem 2.21 in Fan and Yao (2003); Lemma A.1 in Gao (2007); Lemma A.1 in Chen et al. (2012b)) to prove the normality for the second term on RHS of (B.10). Write

$$\sqrt{NT} \cdot Q_3 (Q_1 - Q_2)^- \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T (Z(x_{it}) - q_i) e_{it} = \sum_{t=1}^T V_{NT}(t), \quad (\text{B.11})$$

where $V_{NT}(t) = \frac{1}{\sqrt{NT}} \sum_{i=1}^N Q_3 (Q_1 - Q_2)^- (Z(x_{it}) - q_i) e_{it}$. Observe that Q_3 is such a matrix that selects the first d elements of $\hat{\beta}$, so V_{NT} is a summation of random vectors with finite dimensions $d \times 1$. Then, the conventional Central Limit Theory (CLT) applies. Then partition the set $\{1, \dots, T\}$ into $2\kappa_T + 1$ subsets with large block with size l_T , small block with size s_T and the remaining set with size $T - \kappa_T(l_T + s_T)$, where $l_T = \lfloor T^{(\lambda-1)/\lambda} \rfloor$, $s_T = \lfloor T^{1/\lambda} \rfloor$ and $\kappa_T = \lfloor T/(l_T + s_T) \rfloor$ for any $\lambda > 2$. For $\rho = 1, \dots, \kappa_T$, let $\hat{V} = \sum_{t=\kappa_T(l_T+s_T)+1}^T V_{NT}(t)$, $\tilde{V}_\rho = \sum_{t=\rho(l_T+s_T)+1}^{\rho l_T + (\rho-1)s_T} V_{NT}(t)$ and $\bar{V}_\rho = \sum_{t=\rho l_T + (\rho-1)s_T + 1}^{\rho(l_T+s_T)} V_{NT}(t)$.

For the small blocks, it can be seen that

$$E \left\| \sum_{\rho=1}^{\kappa_T} \bar{V}_\rho \right\|^2 = \sum_{u=1}^d \left\{ \sum_{\rho=1}^{\kappa_T} E[\bar{v}_{\rho,u}^2] + 2 \sum_{\rho=2}^{\kappa_T} (\kappa_T - \rho + 1) E[\bar{v}_{1,u} \bar{v}_{\rho,u}] \right\},$$

where $\bar{V}_\rho = \sum_{t=\rho l_T + (\rho-1)s_T + 1}^{\rho(l_T+s_T)} V_{NT}(t) = (\bar{v}_{\rho,1}, \dots, \bar{v}_{\rho,d})'$. By the properties of α -mixing time series and a procedure similar to (A.6) in Chen et al. (2012b), we obtain $E \left\| \sum_{\rho=1}^{\kappa_T} \bar{V}_\rho \right\|^2 = O \left(\frac{\kappa_T s_T}{T} \right) = o(1)$. Analogously, $E \|\hat{V}\|^2 = O \left(\frac{T - \kappa_T l_T}{T} \right) = o(1)$.

Therefore, in order to establish the CLT, we need only to consider $\sum_{\rho=1}^{\kappa_T} \tilde{V}_\rho$. In connection with Proposition 2.6 in Fan and Yao (2003) and the condition on the α -mixing coefficient, we have

$$\left| E \left[\exp \left\{ \sum_{\rho=1}^{\kappa_T} \|\tilde{V}_\rho\| \right\} \right] - \prod_{\rho=1}^{\kappa_T} E \left[\exp \left\{ \|\tilde{V}_\rho\| \right\} \right] \right| \leq C (\kappa_T - 1) \alpha(s_T) \rightarrow 0,$$

where C is a constant, and $\alpha(\cdot)$ denotes the upper bounded of the α -mixing coefficient provided in Assumption 1 and is achievable in the same way as Assumption A.4 of Chen et al. (2012b).

Then, we know that \tilde{V}_ρ for $\rho = 1, \dots, \kappa_T$ are asymptotically independent. Furthermore, as in the proof of Theorem 2.21.(ii) in Fan and Yao (2003), we have $\text{Cov}[\tilde{V}_1] = \frac{l_T}{T} \Xi_0 (I + o(1))$, where

$$\Xi_0 = \lim_{N, T \rightarrow \infty} \frac{1}{NT} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T Q_3 (Q_1 - Q_2)^- E[e_{it} e_{js} (Z(x_{it}) - q_i) (Z(x_{js}) - q_j)'] (Q'_1 - Q'_2)^- Q'_3.$$

It further implies that $\sum_{\rho=1}^{\kappa_T} \text{Cov}[\tilde{V}_\rho] = \kappa_T \cdot \text{Cov}[\tilde{V}_1] = \frac{\kappa_T l_T}{T} \Xi_0 (I + o(1)) \rightarrow \Xi_0$. Thus, the Feller condition is satisfied.

Moreover, by Cauchy–Schwarz inequality, we have

$$\begin{aligned} E \left[\|\tilde{V}_\rho\|^2 \cdot I \{ \|V_\rho\| \geq \varepsilon \} \right] &\leq \left\{ E \left[\|\tilde{V}_\rho\|^3 \right] \right\}^{2/3} \cdot \left\{ P \left(\|\tilde{V}_\rho\| \geq \varepsilon \right) \right\}^{1/3} \\ &\leq C \left\{ E \left[\|\tilde{V}_\rho\|^3 \right] \right\}^{2/3} \cdot \left\{ E \left[\|\tilde{V}_\rho\|^2 \right] \right\}^{1/3}. \end{aligned}$$

By Lemma B.2 in Chen et al. (2012b),

$$E \left[\|\tilde{V}_\rho\|^3 \right] \leq \left(\frac{l_T}{T} \right)^{3/2} \left\{ E \left[\left\| \frac{1}{\sqrt{N}} \sum_{i=1}^N Q_3 (Q_1 - Q_2)^{-1} (Z(x_{i1}) - q_i) e_{i1} \right\|^4 \right] \right\}^{3/4} = O \left(\left(\frac{l_T}{T} \right)^{3/2} \right),$$

where the last equality follows from the assumption in the body of the theorem. Therefore, we have $E \left[\|\tilde{V}_\rho\|^2 \cdot I \{ \|V_\rho\| \geq \varepsilon \} \right] \leq O \left(\left(\frac{l_T}{T} \right)^{4/3} \right) = o \left(\frac{l_T}{T} \right)$. Consequently,

$$\sum_{\rho=1}^{\kappa_T} E \left[\|\tilde{V}_\rho\|^2 \cdot I \{ \|V_\rho\| \geq \varepsilon \} \right] = o \left(\frac{\kappa_T l_T}{T} \right) = o(1).$$

Then, the Lindeberg condition is satisfied, and therefore the proof is completed. \blacksquare

Proof of Theorem 3.2:

The consistency follows directly from Lemma 3.1. We show next the normality. Using the new selection matrix Q_3 , the asymptotic normality for the term $\sqrt{NT}Q_3(\hat{\beta} - \beta)$ follows in the exact same way as the proof of Theorem 3.1. Hence, by the definition of $\hat{\theta}$ that

$$\sqrt{NT}(\hat{\theta} - \theta_0) = Q_4 \sqrt{NT}Q_3(\hat{\beta} - \beta) + o_P(1) \rightarrow_D N(0, Q\Xi_0Q), \quad (\text{B.12})$$

since $Q_4 \rightarrow_P Q = c_{m_0}^{-1} \theta_{01}^{-m_0+1} \text{diag} \left(1, m_0^{-1/2}, \dots, m_0^{-1/2} \right)$.

Proof of Lemma 4.1:

By (2.5), we have the following decomposition:

$$\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T g(x'_{it}\theta_0) = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it})' \beta + c_0 + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \delta_k(x'_{it}\theta_0).$$

Moreover, $\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T g(x'_{it}\theta_0) = O_P \left(\frac{1}{\sqrt{NT}} \right)$ by the second result of Lemma B.2. Thus,

$$\hat{c}_0 - c_0 = \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it})' (\beta - \hat{\beta}) + \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \delta_k(x'_{it}\theta_0) + O_P \left(\frac{1}{\sqrt{NT}} \right). \quad (\text{B.13})$$

For the first term on RHS of (B.13), we have

$$\left(\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it})' (\beta - \hat{\beta}) \right)^2 \leq \left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it}) \right\|^2 \cdot \|\hat{\beta} - \beta\|^2$$

$$= O_P \left(\frac{k^{3d}}{NT} \right) + O_P \left(k^{-r+3d/2} \right),$$

where $\left\| \frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T Z(x_{it}) \right\| = O_P(\sqrt{k^{3d/2}})$ follows similarly from the derivation of the second term on the RHS of (B.2).

By Cauchy–Schwarz inequality and Assumption 3.ii,

$$E \left[\frac{1}{NT} \sum_{i=1}^N \sum_{t=1}^T \delta_k(x'_{it}\theta_0) \right]^2 \leq \frac{1}{N^2 T^2} \sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^T \sum_{s=1}^T \{E[\delta_k^2(x'_{it}\theta_0)] \cdot E[\delta_k^2(x'_{js}\theta_0)]\}^{1/2} = O(k^{-r}).$$

Based on the above, the result follows. ■

Proof of Lemma 4.2:

By the uniqueness of the Moore–Penrose inverse and the fourth result of Lemma B.3 of this appendix above, for each k , the $(k-1) \times (k-1)$ matrix $\left[\frac{1}{NT} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-}$ is the inverse of $\frac{1}{NT} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta})$. Therefore, we obtain

$$\begin{aligned} \hat{\mathcal{C}} - \mathcal{C} &= \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \left(\mathcal{G}(\theta_0) - \mathcal{G}(\hat{\theta}) \right) \\ &\quad + \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{D}(\hat{\theta}) \\ &\quad + \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{E}, \end{aligned} \tag{B.14}$$

where $\mathcal{G}(\theta)_{NT \times 1} = (g(x'_{11}\theta), \dots, g(x'_{1T}\theta), \dots, g(x'_{N1}\theta), \dots, g(x'_{NT}\theta))'$ for $\forall \theta \in \Theta$.

Similar to (B.6), we have

$$\left\| \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{D}(\hat{\theta}) \right\|^2 = O_P(k^{-r}).$$

By the third and fourth results of Lemma B.3 and the similar procedure of (B.4), we obtain

$$\left\| \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{E} \right\|^2 = O_P \left(\frac{k^{3/2}}{NT} \right).$$

Then, we need only to consider the next term. By the same proof as (B.5) and Assumption 4.iii, we write

$$\begin{aligned} &\left\| \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \left(\mathcal{G}(\theta_0) - \mathcal{G}(\hat{\theta}) \right) \right\|^2 \\ &\leq \left(\lambda_{\min} \left(\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) / (NT) \right) \right)^{-1} \cdot \lambda_{\max}(\tilde{W}) \cdot \left(\left\| \mathcal{G}(\theta_0) - \mathcal{G}(\hat{\theta}) \right\|^2 / (NT) \right) \\ &\leq \left(\lambda_{\min} \left(\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) / (NT) \right) \right)^{-1} \cdot \lambda_{\max}(\tilde{W}) \cdot \left(\|\mathcal{X}\|^2 \cdot \left\| \theta_0 - \hat{\theta} \right\|^2 / (NT) \right), \end{aligned}$$

where $\mathcal{X}_{NT \times 1} = (M(x_{11}), \dots, M(x_{1T}), \dots, M(x_{N1}), \dots, M(x_{NT}))'$ and

$$\tilde{W} = (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \left(\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right)^{-} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}).$$

Since \tilde{W} is symmetric and idempotent, $\lambda_{\max}(\tilde{W}) = 1$. By Assumption 4.iii and Theorem 3.1, we know that $\frac{1}{NT} \|\mathcal{X}\|^2 \cdot \|\theta_0 - \hat{\theta}\|^2 = O_P\left(\frac{1}{NT}\right)$. Hence, similar to (B.6), we have

$$\left\| \left[\mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \mathcal{H}(\hat{\theta}) \right]^{-1} \mathcal{H}(\hat{\theta})' (I_N \otimes M_{i_T}) \left(\mathcal{G}(\theta_0) - \mathcal{G}(\hat{\theta}) \right) \right\|^2 = O_P\left(\frac{1}{NT}\right).$$

Based on the above, by the conditions in the body of Theorem 3.1, $O_P\left(\frac{k^{3/2}}{NT}\right)$ is the leading term. Thus, the result follows. \blacksquare

Proof of Theorem 4.1: See the supplementary document of the paper.

Proof of Theorem 4.2: By the orthogonality, $\int_{\mathbb{R}} [H(w)'(\hat{\mathcal{C}} - \mathcal{C})]^2 \exp(-w^2/2) dw = \|\hat{\mathcal{C}} - \mathcal{C}\|^2$. Thus, we obtain

$$\begin{aligned} \|\tilde{g}(w) - g(w)\|_{L^2}^2 &= \int_{\mathbb{R}} (\tilde{g}(w) - g(w))^2 \cdot \exp(-w^2/2) dw \\ &= \int_{\mathbb{R}} \left[H(w)'(\hat{\mathcal{C}} - \mathcal{C}) + \tilde{c}_0 - c_0 - \delta_k(w) \right]^2 \cdot \exp(-w^2/2) dw \\ &= O\left(\|\hat{\mathcal{C}} - \mathcal{C}\|^2\right) + O(|\tilde{c}_0 - c_0|^2) + O(\|\delta_k(w)\|_{L^2}^2) \\ &= O(|\tilde{c}_0 - c_0|^2) + O(\|\delta_k(w)\|^2) = O_P\left(\frac{k^3}{NT}\right) + O(\|\delta_k(w)\|_{L^2}^2), \end{aligned}$$

due to Lemma 4.2 and by the same procedure as Lemma 4.1, $|\tilde{c}_0 - c_0|^2 = O_P\left(\frac{k^3}{NT}\right)$. The result follows. \blacksquare

References

- D. W. K. Andrews. Cross-section regression with common shocks. *Econometrica*, pages 1551–1585, 2005.
- J. Bai. Panel data models with interactive fixed effects. *Econometrica*, pages 1229–1279, 2009.
- B. H. Baltagi, J. M. Griffin, and W. Xiong. To pool or not to pool: Homogenous versus heterogenous estimators applied to cigarette demand. *Reviews of Economics and Statistics*, pages 117–126, 2000.
- G. Blower. *Random Matrices: High Dimensional Phenomena*. Cambridge University Press, New York, 2009.
- R. Carroll, J. Fan, I. Gijbels, and M. P. Wand. Generalized partially linear single-index models. *Journal of the American Statistical Association*, pages 477–489, 1997.
- J. Chen, J. Gao, and D. Li. Semiparametric trending panel data models with cross-sectional dependence. *Journal of Econometrics*, pages 71–85, 2012a.
- J. Chen, J. Gao, and D. Li. A new diagnostic test for cross-section uncorrelatedness in nonparametric panel data models. *Econometric Theory*, pages 1–20, 2012b.
- J. Chen, J. Gao, and D. Li. Estimation in single-index panel data models with heterogeneous link functions. *Econometric Review*, pages 928–955, 2013a.
- J. Chen, J. Gao, and D. Li. Estimation in partially linear single-index panel data models with fixed effects. *Journal of Business and Economic Statistics*, pages 315–330, 2013b.

- X. Chen. Large sample sieve estimation of semi-nonparametric models. *Handbook of Econometrics*, pages 5549–5632, 2007.
- C. Dong and J. Gao. Specification testing in structural nonparametric cointegration. *Working Paper 02/14, Department of Econometrics and Business Statistics, Monash Univeristy*, 2014.
- C. Dong, J. Gao, and D. Tjøstheim. Estimation for single-index and partially linear single-index nonstationary time series models. *Working Paper 07/14, Department of Econometrics and Business Statistics, Monash Univeristy*, 2014.
- J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods*. Springer-Verlag, 2003.
- A. R. Gallant and D. W. Nychka. Semi-nonparametric maximum likelihood estimation. *Econometrica*, pages 363–390, 1987.
- A. R. Gallant and G. Tauchen. Semiparametric estimation of conditionally constrained heterogeneous processes: Asset pricing applications. *Econometrica*, pages 1091–1120, 1989.
- J. Gao. *Nonlinear Time Series: Semiparametric and Nonparametric Methods*. Chapman and Hall, 2007.
- J. Gao and P.C.B. Phillips. Semiparametric estimation in triangular system equations with nonstationarity. *Journal of Econometrics*, pages 59–79, 2013.
- J. Gao, H. Tong, and R. Wolff. Model specification tests in nonparametric stochastic regression models. *Journal of Multivariate Analysis*, pages 324–359, 2002.
- J. L. Horowitz. *Semiparametric and Nonparametric Methods in Econometrics*. Springer Science+Business Media, LLC, 2009.
- H. Ichimura. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *Journal of Econometrics*, pages 71–120, 1993.
- J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons Ltd, third edition, 2007.
- P. Nevai. Géza fredu, orthogonal polynomials and christoffel functions. A case study. *Journal of Approximation Theory*, pages 3–167, 1986.
- W. K. Newey. Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, pages 147–168, 1997.
- W. K. Newey and J. L. Powell. Instrumental variable estimation of nonparametric models. *Econometrica*, pages 1565–1578, 2003.
- M. H. Pesaran. Estimation and inference in large heterogeneous panels with a multifactor error structure. *Econometrica*, pages 967–1012, 2006.
- L. Su and S. Jin. Sieve estimation of panel data models with cross section dependence. *Journal of Econometrics*, pages 34–47, 2012.
- Q. Wang and P.C.B. Phillips. Structural nonparametric cointegrating regression. *Econometrica*, pages 1901–1948, 2009.