ISSN 1440-771X

MONASH University

Australia

Department of Econometrics and Business Statistics

http://www.buseco.monash.edu.au/depts/ebs/pubs/wpapers/

Semiparametric Single-Index Panel Data Models With Cross-Sectional Dependence

Bin Peng, Chaohua Dong and Jiti Gao

February 2014

Working Paper 09/14

Semiparametric Single–Index Panel Data Models with Cross–Sectional Dependence[1]

Bin Peng[2]

Economics Discipline Group

University of Technology Sydney

Chaohua Dong and Jiti Gao

Department of Econometrics and Business Statistics

Monash University

Abstract

In this paper, we consider a semiparametric single index panel data model with cross–sectional dependence, high–dimensionality 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 a link function for the case where both N and T go to ∞. Rates of convergence and asymptotic normality consistencies 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 method.

JEL classification: C13, C14, C23.

[1]The authors of this paper acknowledge the Australian Research Council Discovery Grants Program support under Grant numbers: DP1096374 and DP130104229.

[2]Address for correspondence: Bin Peng, Economics Discipline Group, University of Technology Sydney, Australia. Email: pengbin430@gmail.com.

1 Introduction

Single-index models have been studied by both econometricians and statisticians in the past twenty years. These models include many classic parametric models (e.g. linear model or logistic model) by using a general function form $g(x'\beta)$ (see, for example, Chapter 2 of Gao (2007)). For nonlinear panel data models, the researcher starts introducing single–index panel data models (cf. 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 the 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 convergence rate $\sqrt{NT}$, 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. Meanwhile, Su and Jin (2012) propose using sieve estimation techniques to a nonparmetric multi–factor model, which is a nonparametric version of the parametric counterpart proposed in Pesaran (2006).

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 ∞. The estimation procedure proposed below allows us to avoid some computational issues and is therefore easy to implement. In this paper, we consider the stationary time series case. Non-stationary situations are much more complex and will be discussed in a companion paper. The estimation techniques proposed in this paper can also be extended to the multi–factor structure model. (Under certain restrictions similar to those of Su and Jin (2012), a semiparametric single–index extension can be achieved.) Furthermore, we add fixed effects to the model and do not impose any particular assumptions on them, so 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. Moreover, we avoid the issue about the curse of dimensionality through using a single–index form for the regressors.

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 (cf. 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, the 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.

In summary, this paper makes the following contributions:

1. It proposes a semiparametric single–index panel data model to simultaneously accommodate cross–sectional dependence, high–dimensionality, stationarity and unobservable heterogeneity;

2. It establishes simple and consistent closed–form estimates for both unknown parameters and link function, and the closed–form estimates are easy to implement in practice;

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 and through using both simulated and real data examples.

The structure of this paper is as follows. Section 2 introduces 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 provide a simple Monte Carlo experiment and two empirical case studies by looking into UK's climate data and US cigarettes demand. Section 6 concludes this paper with some comments. All the proofs are given in an appendix.

Throughout the paper, we will use the following notation: $\otimes$ denotes the Kronecker product; $vec(A)$ defines the vec operator that transforms a matrix A into a vector by stacking the columns of the matrix one underneath the other; I_k denotes an identity matrix with dimensions $k \times k$; i_k denotes a $k \times 1$ one vector $(1, \ldots, 1)'$; $M_p = I_k - P(P'P)^{-1}P'$ denotes the project matrix generated by matrix P with dimensions $k \times h$ and $h \leq k$; A^- denotes the Moore-Penrose inverse of the matrix A; $\xrightarrow{P}$ denotes converging in probability; $\xrightarrow{D}$ denotes converging in distribution; $\|\cdot\|$ denots the Euclidean norm; $\lfloor a \rfloor \leq a$ means the largest integer part of a.

2 Semiparametric Single-Index Panel Data Models

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

$$y_{it} = g\left(x_{it}'\theta_0\right) + \gamma_i + e_{it}, \quad i = 1, \ldots, N, \quad t = 1, \ldots, T, \tag{2.1}$$

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 $g : \mathbb{R} \to \mathbb{R}$ is unknown. We use γ_i's to capture fixed effects in this model, which are allowed to be correlated with the regressors. Under the current set-up, our main interests are to consistently estimate the vector of unknown parameters $\theta_0 = (\theta_{01}, \ldots, \theta_{0d})'$ and link function $g(\cdot)$ for the case where both N and T go to ∞.

To ensure that identification requirements are satisfied (cf. Horowitz (2009) and Ichimura (1993)), we assume that θ_0 belongs to a compact set Θ, $\|\theta_0\| = 1$ and $\theta_{01} > 0$. For the link function $g(\cdot)$, we expand it by Hermite polynomials and approximate it by a linear combination of a finite number of basis functions from the expansion. As the number of basis functions increases, the proxy approaches the true function. By doing so, a nonparametric estimation is practically turned to a parametric one, so we need only to estimate θ_0 and the coefficients of the basis functions simultaneously.

We now introduce the background of Hermite polynomials briefly and explain how to expand the link function. Hermite polynomial system $\{H_m(w),\ m = 0, 1, 2, \ldots\}$ is a complete orthogonal system in a Hilbert space $L^2\left(\mathbb{R}, \exp\left(-w^2/2\right)\right)$ and each element is denoted as

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

Since $\int_{\mathbb{R}} H_m(w) H_n(w) \exp\left(-w^2/2\right) dw$ equals to $m!\sqrt{2\pi}$ for $m = n$ and 0 for $m \neq n$ respectively, the normalised orthogonal system is denoted as $\{h_m(w),\ m = 0, 1, 2, \ldots\}$, where $h_m(w) = \frac{1}{\sqrt{m!\sqrt{2\pi}}} H_m(w)$.

Thus, for $\forall g \in L^2\left(\mathbb{R}, \exp\left(-w^2/2\right)\right)$, we can express it in terms of $h_m(w)$ as follows:

$$g(w) = \sum_{m=0}^{\infty} c_m h_m(w) \text{ and } c_m = \int_{\mathbb{R}} g(w) \cdot h_m(w) \cdot \exp\left(-w^2/2\right) dw. \tag{2.3}$$

Furthermore, $h_m(w) \cdot \exp(-w^2/4)$ is bounded uniformly in $w \in \mathbb{R}$ and m (cf. Nevai (1986)).

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. Since neither of these two methods results in a closed form estimation method, numerical estimates are often sensitive to the initial values used in practice due to multi–modality or flatness of a curve. Instead, we further expand $h\left(x_{it}'\theta_0\right)$ by Lemma 1 of the appendix as follows:

$$g\left(x_{it}'\theta_0\right) = \sum_{m=0}^{k-1} c_m h_m\left(x_{it}'\theta_0\right) + \sum_{m=k}^{\infty} c_m h_m\left(x_{it}'\theta_0\right) \tag{2.4}$$

$$= \sum_{m=0}^{k-1} \sum_{|p|=m} a_{mp}\left(\theta_0\right) \mathscr{H}_p\left(x_{it}\right) + \delta_k\left(x_{it}'\theta_0\right), \tag{2.5}$$

where

$$\delta_k\left(x_{it}'\theta_0\right) = \sum_{m=k}^{\infty} c_m h_m\left(x_{it}'\theta_0\right), \quad a_{mp} = \binom{m}{p} c_m \theta_0^p, \quad \binom{m}{p} = \frac{m!}{\prod_{j=1}^{d} p_j!},$$

$$\theta_0^p = \prod_{j=1}^{d} \theta_{0j}^{p_j}, \quad \mathscr{H}_p\left(x_{it}\right) = \prod_{j=1}^{d} h_{p_j}\left(x_{it,j}\right), \quad x_{it} = \left(x_{it,1}, \ldots, x_{it,d}\right)', \quad p = \left(p_1, \ldots, p_d\right)',$$

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

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 $\delta_k\left(x_{it}'\theta_0\right)$ can be considered as a truncated error term, which goes to zero as k increases. Since each $h_m(w) \cdot \exp(-w^2/4)$ is bounded uniformly in $w \in \mathbb{R}$ and m, $\mathscr{H}_p(x) \cdot \exp(-\|x\|^2/4)$ must be bounded uniformly in $x \in \mathbb{R}^d$ and p.

To further investigate the model, we now define an ordering relationship with respect to p in (2.5).

Definition 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 $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_d) < \check{p} = (\check{p}_1, \ldots, \check{p}_d)$ if $\hat{p}_j = \check{p}_j$ for all $j = 1, \ldots, l-1$ and $\hat{p}_l < \check{p}_l$, where $1 < l \leq d$.*

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

- As $m = 0$,

$$p = (0, 0, \ldots, 0)', \quad \mathscr{H}_p\left(x_{it}\right) = h_0\left(x_{it,1}\right) h_0\left(x_{it,2}\right) \cdots h_0\left(x_{it,d}\right) = 1,$$
$$a_{0p}(\theta_0) = c_0. \tag{2.6}$$

- As $m = 1$,

$$\begin{aligned}
p = (1, 0, \ldots, 0)', \quad \mathscr{H}_p(x_{it}) &= h_1(x_{it,1}) h_0(x_{it,2}) \cdots h_0(x_{it,d}) = x_{it,1}, \\
a_{1p}(\theta_0) &= c_1 \theta_{01}; \\
p = (0, 1, \ldots, 0)', \quad \mathscr{H}_p(x_{it}) &= h_0(x_{it,1}) h_1(x_{it,2}) \cdots h_0(x_{it,d}) = x_{it,2}, \\
a_{1p}(\theta_0) &= c_1 \theta_{02}; \\
&\vdots \\
p = (0, 0, \ldots, 1)', \quad \mathscr{H}_p(x_{it}) &= h_0(x_{it,1}) h_0(x_{it,2}) \cdots h_1(x_{it,d}) = x_{it,d}, \\
a_{1p}(\theta_0) &= c_1 \theta_{0d}. \qquad (2.7)
\end{aligned}$$

- As $m = 2, \ldots, k-1$,

$$\begin{aligned}
p = (m, 0, \ldots, 0)', \quad \mathscr{H}_p(x_{it}) &= h_m(x_{it,1}) h_0(x_{it,2}) \cdots h_0(x_{it,d}) = h_m(x_{it,1}), \\
a_{mp}(\theta_0) &= c_m \theta_{01}^m; \\
p = (m-1, 1, 0, \ldots, 0)', \quad \mathscr{H}_p(x_{it}) &= h_{m-1}(x_{it,1}) h_1(x_{it,2}) h_0(x_{it,3}) \cdots h_0(x_{it,d}) \\
&= h_{m-1}(x_{it,1}) h_1(x_{it,2}), \\
a_{mp}(\theta_0) &= m c_m \theta_{01}^{m-1} \theta_{02}; \\
&\vdots \\
p = (0, 0, \ldots, m)', \quad \mathscr{H}_p(x_{it}) &= h_0(x_{it,1}) h_0(x_{it,2}) \cdots h_m(x_{it,d}) = h_m(x_{it,d}), \\
a_{mp}(\theta_0) &= c_m \theta_{0d}^m. \qquad (2.8)
\end{aligned}$$

Note that, by (2.6), it is easy to know that the first $\mathscr{H}_p(x_{it})$ in (2.5) is constant one and its coefficient is constant c_0. The second to the $(d+1)^{\text{th}}$ $\mathscr{H}_p(x_{it})$'s in (2.5) are simply x_{it} with coefficients $c_1\theta_0$ shown in (2.7) and will be used to recover the interest parameter θ_0 later on.

Accordingly, it allows us to denote the next two vectors to shorten notation:

$$Z(x_{it}) = \left(Z_1(x_{it})', \ldots, Z_{k-1}(x_{it})'\right)' \quad \text{and} \quad \beta = \left(A_1(\theta_0)', \ldots, A_{k-1}(\theta_0)'\right)',$$

where, for $m = 1, \ldots, k-1$, $Z_m(x_{it})$'s are column vectors consisting of all $\mathscr{H}_p(x_{it})$'s arranged on descent ordering with respect to $|p| = m$ and $A_m(\theta_0)$ are column vectors consisting of all corresponding $a_{mp}(\theta_0)$. Notice that we have suppressed the notation θ_0 in β for simplicity. Thus, we can rewrite the model (2.1) as

$$\begin{aligned}
y_{it} &= c_0 + H(x_{it}'\theta_0)' \mathcal{C} + \delta_k(x_{it}'\theta_0) + \gamma_i + e_{it} \\
&= c_0 + Z(x_{it})' \beta + \delta_k(x_{it}'\theta_0) + \gamma_i + e_{it},
\end{aligned}$$

where $c_0 = a_{0p}(\theta_0)$, $\mathcal{C} = (c_1, c_2, \ldots, c_{k-1})'$ and $H(w) = (h_1(w), h_2(w), \ldots, h_{k-1}(w))'$ for $w \in \mathbb{R}$.

Moreover, it is easy to check that the cardinality of P_m is $\binom{m+d-1}{d-1}$, so the length of the vector $Z(x_{it})$ is

$$K = \sum_{m=1}^{k-1} \# P_m = \frac{(d+k-1)!}{d!\,(k-1)!} - 1 = O\left(k^d\right). \tag{2.9}$$

Then, we may write $Z(x_{it})$ as

$$Z(x_{it}) = \left(Z_1(x_{it})', \ldots, Z_{k-1}(x_{it})'\right)' = (z_1(x_{it}), \ldots, z_K(x_{it}))',$$

where $z_u(x_{it})$'s for $u = 1, \ldots, K$ are $\mathscr{H}_p(x_{it})$'s in (2.7) and (2.8) in the exactly same order.

To remove the fixed effects, we introduce the following notation:

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

and then eliminate γ_i's by the within–transformation. The model now becomes

$$\begin{aligned}\tilde{y}_{it} &= \tilde{H}_{it}(\theta_0)'\mathcal{C} + \tilde{\delta}_k(x_{it}'\theta_0) + \tilde{e}_{it} \\ &= \tilde{Z}_{it}'\beta + \tilde{\delta}_k(x_{it}'\theta_0) + \tilde{e}_{it}.\end{aligned}$$

Alternatively, we can express the model in matrix forms as

$$(I_N \otimes M_{i_T})\,\mathcal{Y} = (I_N \otimes M_{i_T})\,\mathcal{H}(\theta_0)\mathcal{C} + (I_N \otimes M_{i_T})\,\mathcal{D}(\theta_0) + (I_N \otimes M_{i_T})\,\mathcal{E} \tag{2.10}$$
$$= (I_N \otimes M_{i_T})\,\mathcal{Z}\beta + (I_N \otimes M_{i_T})\,\mathcal{D}(\theta_0) + (I_N \otimes M_{i_T})\,\mathcal{E}, \tag{2.11}$$

where

$$\underset{NT\times 1}{\mathcal{Y}} = (y_{11}, \ldots, y_{1T}, \ldots, y_{N1}, \ldots, y_{NT})',$$
$$\underset{NT\times(k-1)}{\mathcal{H}(\theta)} = (H(x_{11}'\theta), \ldots, H(x_{1T}'\theta), \ldots, H(x_{N1}'\theta), \ldots, H(x_{NT}'\theta))' \text{ for } \forall\theta \in \Theta,$$
$$\underset{NT\times 1}{\mathcal{D}(\theta)} = (\delta_k(x_{11}'\theta), \ldots, \delta_k(x_{1T}'\theta), \ldots, \delta_k(x_{N1}'\theta), \ldots, \delta_k(x_{NT}'\theta))' \text{ for } \forall\theta \in \Theta,$$
$$\underset{NT\times K}{\mathcal{Z}} = (Z(x_{11}), \ldots, Z(x_{1T}), \ldots, Z(x_{N1}), \ldots, Z(x_{NT}))',$$
$$\underset{NT\times 1}{\mathcal{E}} = (e_{11}, \ldots, e_{1T}, \ldots, e_{N1}, \ldots, e_{NT})'.$$

Notice that c_0 is a constant, so it is also removed by the within–transformation. It indicates that one can only identify the unknown function $g(\cdot)$ up to a constant through (2.10)-(2.11). To estimate the location, extra assumptions are needed (e.g. Assumption 1.ix in Su and Jin (2012)). In the next section, we will recover the interest parameter θ_0 by (2.11). After that, we will bring a consistent estimate for θ_0 back to (2.10) and recover the link function in section 4.

3 Estimation of Parameter θ_0

We consider a within–ordinary least squares (OLS) estimator of β:

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

To simplify the notation, for each time series $\{x_{i1}, \ldots, x_{iT}\}$, let $Q_{1,i} = E\left[Z(x_{it}) Z(x_{it})'\right]$ and $q_i = E[Z(x_{it})]$. Also, denote 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 $t = 1, \ldots, T$, let $x_t = (x_{1t}, \ldots, x_{Nt})'$ and $e_t = (e_{1t}, \ldots, e_{Nt})'$.

We now are ready to introduce the following assumptions. Specifically, we do not impose any assumption on the fixed effects in this paper, so they can be correlated with the regressors to capture unobservable heterogeneity.

Assumption 1 (Covariates and errors):

i. Let $E[e_{it}] = 0$ for all $i \geq 1$ and $t \geq 1$. Suppose that $\{x_t, e_t : t \geq 1\}$ is strictly stationary and α-mixing. Let $\alpha_{ij}(|t-s|)$ represent the α-mixing coefficient between $\{x_{it}, e_{it}\}$ and $\{x_{js}, e_{js}\}$. Let the α-mixing coefficients satisfy

$$\sum_{i=1}^N \sum_{j=1}^N \sum_{t=1}^\infty (\alpha_{ij}(t))^{\eta/(4+\eta)} = O(N) \text{ and } \sum_{i=1}^N \sum_{j=1}^N (\alpha_{ij}(0))^{\eta/(4+\eta)} = O(N),$$

where $\eta > 0$ is chosen such that $E\left[|e_{it}|^{4+\eta}\right] < \infty$ and $E\left[\|x_{it}\|^{4+\eta}\right] < \infty$.

ii. Suppose that x_{it} is independent of e_{js} for all $i, j \geq 1$ and $t, s \geq 1$.

Assumption 1.i entails that only the stationary cases are considered in this paper. The nonstationary cases are more complex and will be discussed in a companion paper. 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 spirit the same 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 by doing so. Thus, all the cross-sectional dependences and time series properties are captured by the mixing coefficients. Particularly, $\alpha_{ij}(0)$ only measures the cross-sectional dependence between $\{x_{it}, e_{it}\}$ and $\{x_{jt}, e_{jt}\}$.

We now use the factor model structure as an example to show that Assumption 1.i is verifiable. Suppose that $e_{it} = \gamma_i f_t + \varepsilon_{it}$, where all variables are scalars and ε_{it} is independent and identically distributed (i.i.d.) across i and t with mean zero. Simple algebra shows that the coefficient $\alpha_{ij}(|t-s|)$ reduces to $\alpha_{ij} \cdot b(|t-s|)$, in which $\alpha_{ij} = E[\gamma_i \gamma_j]$ and $b(|t-s|)$ is the α–mixing coefficient of the factor time series $\{f_1, \ldots, f_T\}$. If f_t is strictly stationary α-mixing process and γ_i is i.i.d. or a m-dependent sequence (cf. Appendix A of Gao (2007)Definition 9.1 and Theorem 9.1 in DasGupta (2008)), Assumption 1.i can easily be verified. More details and useful empirical examples can be found under Assumption A2 in Chen et al. (2012a).

Assumption 2 (Identifications):

i. Let Θ be a compact subset of $\mathbb{R}^d$ and $\theta_0 \in \Theta$ be in the interior of Θ. Moreover, $\|\theta_0\| = 1$ and $\theta_{01} > 0$.

ii. $E[g(x'_{it}\theta_0)] = 0$ for all $i \geq 1$ and $t \geq 1$. Moreover, for the same η in Assumption 1, let $E[|g(x'_{it}\theta_0)|^{2+\eta/2}] < \infty$.

Assumption 2.i is a standard identification requirement. Instead one can follow Ichimura (1993) to assume $\theta_{01} = 1$. However, by doing so, it seems to be hard to decide which variable should be considered as constant one in practice. Assumption 2.ii is not really necessary when the main interests are only 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 is estimating $g(x'_{it}\theta_0) - E[g(x'_{it}\theta_0)]$ rather than the true $g(x'_{it}\theta_0)$ (cf. Su and Jin (2012)). An example is given in a Monte Carlo study for illustration.

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.

ii. (1) There exists $r > 0$ such that $\sup_{0 \leq \epsilon \leq 1} \sup_{\|\theta - \theta_0\| \leq \epsilon} \max_{i \geq 1} E[\delta_k^2(x'_{i1}\theta)] = o(k^{-r})$.

(2) $\max_{i \geq 1} E\left[|\mathscr{H}_p(x_{i1})|^{4+\eta}\right] = O\left(|p|^d\right)$ as $|p| \to \infty$, where $|p|$ is given under (2.5).

iii. $\frac{k^{3d}}{NT} \to 0$ and $\frac{k^{3d/2}}{T} \to 0$ as $N, T, k \to \infty$ jointly.

Assumption 3.i can be verified by carrying on a similar procedure 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) and the second condition of this assumption is more general than Assumption 3.iv in Su and Jin (2012). By the argument under Assumption 2 in Newey (1997), it is not possible to assume $\mathscr{H}_p(x_{i1})$ is bounded uniformly. Therefore, we put restrictions on the moments of the basis functions. Compared to putting the bounds on the basis power series directly of Newey (1997), we believe our current assumption is more realistic. Also, the second condition in our Assumption 3.ii clearly allows x_{it} to follow the standard multivariate normal distribution, which is ruled out by Assumption 3.iv in Su and Jin (2012) for the cases where the basis functions are the hermite polynomials. More relevant discussions will be given under Theorem 1.

We now illustrate that it is verifiable by the following example:

Suppose that we consider the second condition in Assumption 3.ii and $\{x_{i1}, \ldots, x_{iT}\}$ has the standard multivariate normal density for $i \geq 1$. Let η be large enough (say $\eta = 1$ without losing generality) and $x = (x_1, \ldots, x_d)$. Then

$$E\left[|\mathscr{H}_p(x_{i1})|^5\right] = \int_{\mathbb{R}^d} |\mathscr{H}_p(x)|^5 \cdot \exp(-\|x\|^2/2)dx.$$

Note $|\mathscr{H}_p(x)|^5 = |h_{p_1}(x_1)\cdots h_{p_d}(x_d)|^5$, so expand it as $\left|\mathscr{H}_p^5(x)\right| = \left|\prod_{j=1}^d \sum_{s_j=0}^{5p_j} b_{s_j} h_{s_j}(x_j)\right|$, which gives that

$$\begin{aligned}\int_{\mathbb{R}} |\mathscr{H}_p(x)|^5 \cdot \exp(-\|x\|^2/2)dx &\leq \prod_{j=1}^d \sum_{s_j=0}^{5p_j} |b_{s_j}| \int_{\mathbb{R}} |h_{s_j}(x_j)| \exp\left(-x_j^2/4\right) \cdot \exp\left(-x_j^2/4\right) dx \\ &\leq \prod_{j=1}^d C_1 \sum_{s_j=0}^{5p_j} |b_{s_j}| \int_{\mathbb{R}} \exp\left(-x_j^2/4\right) dx \\ &\leq \prod_{j=1}^d C_2 \sum_{s_j=0}^{5p_j} |b_{s_j}| \leq C_3 \prod_{j=1}^d 5p_j \leq C_4 |p|^d,\end{aligned}$$

where we have used that $h_{s_j}(x_j)\exp\left(-x_j^2/4\right)$ is bounded uniformly in s_j and x_j, and $|b_{s_j}|$ is bounded over s_j. Then, by moments monotonicity, the second condition in Assumption 3.ii has been verified. Analogously, we can show that the condition 1 in Assumption 3.ii is verifiable.

Assumption 3.iii implies that the rate of $k \to \infty$ needs to be slower than that of $\min\{(NT)^{\frac{1}{3d}}, T^{\frac{2}{3d}}\}$. In practice, the lengths of the cross-sectional dimension and time series can be relatively large, so Assumption 3.iii is easy to achieve. Moreover, the researcher normally assumes that $N/T \to c \in (0, \infty]$ as $N, T \to \infty$ in the conventional panel data case, which is also covered by Assumption 3.iii.

We are ready to establish the main results and their proofs are given in the appendix.

Theorem 1 *Let Assumptions 1, 2.i and 3 hold. Then, we have*

$$\left\|\hat{\beta} - \beta\right\|^2 = O_p\left(\frac{k^{3d/2}}{NT}\right) + o_p\left(k^{-r}\right).$$

The first term of the convergence rate is not the optimal rate $O_p\left(\frac{k^d}{NT}\right)$, which is due to the fact that we can not bound the hermite polynomials uniformly. However, the optimality is achievable when the fourth order moment is bounded uniformly. This may be done in the same way as in Su and Jin (2012). By doing so, we will rule out the standard multivariate normal density for x_{it} at least. The same arguments also apply to the other convergency rates given below.

Notice that the first d elements of β only involve θ_0 and constant c_1 by (2.7). Moreover, $\left\|\hat{\beta}^d - \beta^d\right\|^2 \leq \left\|\hat{\beta} - \beta\right\|^2$, where $\hat{\beta}^d$ and β^d denote the first d elements of $\hat{\beta}$ and β, respectively. In connection with the identification restriction, it is easy to obtain that $\sqrt{\sum_{i=1}^d \hat{\beta}_i^2}$ converges to $|c_1|$. Then, intuitively, the estimator of θ_0 is as follows.

$$\hat{\theta} = \frac{sgn(\hat{\beta}_1)}{\sqrt{\sum_{i=1}^d \hat{\beta}_i^2}} \cdot Q_3 \cdot \hat{\beta}, \quad \underset{d\times K}{Q_3} = \begin{pmatrix} I_d & \underset{d\times(K-d)}{0} \end{pmatrix} \text{ and } I_d \text{ is a } d\times d \text{ identity matrix.}$$

By Theorem 1, the following corollary follows immediately.

Corollary 1 *Under the conditions of Theorem 1, $\hat{\theta}$ is consistent.*

Furthermore, we establish the following normality.

Theorem 2 *Let Assumptions 1, 2.i and 3 hold. If, in addition, $\frac{NT}{k^r} \to \sigma$ for $0 \leq \sigma < \infty$, $\frac{k^{4.5d}}{NT} \to 0$ and $E\left[\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^N Q_3\left(Q_1 - Q_2\right)^{-1}\left(Z\left(x_{i1}\right) - q_i\right)e_{i1}\right\|^4\right] = O(1)$, as $(N,T) \to (\infty, \infty)$ jointly, then*

$$\sqrt{NT}\left(\hat{\theta} - \theta_0\right) \xrightarrow{D} N\left(0, c_1^{-2}\cdot\Xi_0\right),$$

where

$$\Xi_0 = \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} Q_3\left(Q_1 - Q_2\right)^{-}\Big\{E\left[e_{i1}^2\left(Z\left(x_{i1}\right)-q_i\right)\left(Z\left(x_{i1}\right)-q_i\right)'\right]$$
$$+\sum_{t=2}^{\infty} E\left[e_{i1}e_{it}\left(Z\left(x_{i1}\right)-q_i\right)\left(Z\left(x_{it}\right)-q_i\right)'\right]$$
$$+\sum_{t=2}^{\infty} E\left[e_{i1}e_{it}\left(Z\left(x_{it}\right)-q_i\right)\left(Z\left(x_{i1}\right)-q_i\right)'\right]\Big\}\left(Q_1'-Q_2'\right)^{-}Q_3',$$

and c_1 is denoted in (2.3).

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 on $g(x_{it}'\theta_0)$ (see (2.5) for details) and can be considered as a trade-off in order to achieve the closed form estimator.

The restriction $E\left[\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Q_3\left(Q_1-Q_2\right)^{-1}\left(Z\left(x_{i1}\right)-q_i\right)e_{i1}\right\|^4\right] = O(1)$ is in spirit the same as Assumption ii of Lemma A.1 in Chen et al. (2012b) and can be easily verified for the i.i.d cases.

Based on Theorems 1 and 2, it is easy to realise that

$$\hat{\Xi}_0 = Q_3\hat{Q}_{12}^{-1}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\left(\tilde{y}_{it}-\tilde{Z}_{it}'\hat{\beta}\right)^2\right)\hat{Q}_{12}^{-1}Q_3' \xrightarrow{P} \Xi_0,$$

where $\hat{Q}_{12} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'$. Therefore, the traditional hypothesis tests on θ_0 can be established by

$$\left(\sum_{i=1}^{d}\hat{\beta}_i^2\right)^{1/2}\hat{\Xi}_0^{-1/2}\cdot\sqrt{NT}\left(\hat{\theta}-\theta_0\right)\xrightarrow{D} N\left(0, I_d\right).$$

So far we have fully recovered the interest parameter θ_0. We will focus on the link function in the following section.

4 Estimation of The Link Function

We now can only estimate $g(x'\theta_0)$ up to a constant by using $\hat{\beta}$, because c_0 gets cancelled out by the within–transformation. Therefore, we need to take into account the location of the link function by Assumption 2.ii and recover c_0 by the next theorem. The proofs of the following results are given in the appendix.

Theorem 3 *Under Assumptions 1–3, we have*

$$\left(\hat{c}_0 - c_0\right)^2 = O_p\left(\frac{k^{3d/2}}{NT}\right) + o_p\left(k^{-r}\right),$$

where $\hat{c}_0 = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)'\hat{\beta}$.

In connection with (2.5) and Theorem 1, intuitively, we provide the next estimator for $g(x'\theta_0)$.

$$\hat{g}\left(x'\hat{\theta}\right) = Z\left(x\right)'\hat{\beta} + \hat{c}_0 \tag{4.1}$$

Based on the proof of Theorem 2, achieving the next result is straightforward.

Theorem 4 *Let Assumptions 1, 2 and 3 hold. If, in addition,* $\frac{NT}{k^r} \to \sigma$ *for* $r > 4.5d$ *and* $0 \leq \sigma < \infty$, $\frac{k^{4.5d}}{NT} \to 0$ *and* $E\left[\left\|\frac{1}{\sqrt{NK^{3/2}}}\sum_{i=1}^{N}\left(Z\left(x\right)-\bar{q}\right)'\left(Q_1-Q_2\right)^{-1}\left(Z\left(x_{i1}\right)-q_i\right)e_{i1}\right\|^4\right] = O(1)$*, as* $(N,T) \to (\infty,\infty)$ *jointly, then*

$$\sqrt{\frac{NT}{K^{3/2}}}\left(\hat{g}\left(x'\hat{\theta}\right) - g(x'\theta_0)\right) \xrightarrow{D} N\left(0, \Xi_1\right),$$

where

$$\begin{aligned}\Xi_1 = \lim_{N,k\to\infty}\frac{1}{N}\sum_{i=1}^{N}\frac{1}{K^{3/2}}\left(Z\left(x\right)-\bar{q}\right)'\left(Q_1-Q_2\right)^{-}\Bigg\{&E\left[e_{i1}^2\left(Z\left(x_{i1}\right)-q_i\right)\left(Z\left(x_{i1}\right)-q_i\right)'\right]\\ +\sum_{t=2}^{\infty}E\left[e_{i1}e_{it}\left(Z\left(x_{i1}\right)-q_i\right)\left(Z\left(x_{it}\right)-q_i\right)'\right]&\\ +\sum_{t=2}^{\infty}E\left[e_{i1}e_{it}\left(Z\left(x_{it}\right)-q_i\right)\left(Z\left(x_{i1}\right)-q_i\right)'\right]\Bigg\}&\left(Q_1'-Q_2'\right)^{-}\left(Z\left(x\right)-\bar{q}\right),\end{aligned}$$

and K *is chosen by (2.9).*

Again, it is pointed out that while the rate of convergence may not be optimal, the optimality is achievable when the fourth order moment is bounded uniformly. This may be done in the same way as in Su and Jin (2012). However, the optimality comes with a price. For example, x_{it} cannot even follow the standard multivariate normal distribution.

Similar to Theorem 2, it is easy to establish a standardised version of the form:

$$\sqrt{\frac{NT}{K^{3/2}}}\cdot\hat{\Xi}_1^{-1/2}\cdot\left(\hat{g}\left(x'\hat{\theta}\right)-g(x'\theta_0)\right) \xrightarrow{D} N\left(0,1\right),$$

where

$$\hat{\Xi}_1 = (Z(x) - \hat{q})' \hat{Q}_{12}^{-1} \left(\frac{1}{NTK} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Z}_{it}' \left(\tilde{y}_{it} - \tilde{Z}_{it}' \hat{\beta} \right)^2 \right) \hat{Q}_{12}^{-1} (Z(x) - \hat{q}),$$

$$\hat{Q}_{12} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \tilde{Z}_{it} \tilde{Z}_{it}' \quad \text{and} \quad \hat{q} = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} Z(x_{it}).$$

In practice, the above results are useful to calculate the confidence interval for a point prediction of $g(x'\theta_0)$.

Notice that the above two theorems just recover $g(x'\theta_0)$ rather than $g(w)$ itself. To estimate the link function $g(w)$ regardless of θ_0, we now bring $\hat{\theta}$ in (2.10) and then provide our estimator on $\mathcal{C}$ below.

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

We will proceed as in the previous section to investigate (4.2). To simplify the notation, for each time series $\{x_{i1}, \ldots, x_{iT}\}$, let $R_{1,i}(\theta) = E\left[H(x_{it}'\theta) H(x_{it}'\theta)'\right]$ and $r_i(\theta) = E\left[H(x_{it}'\theta)\right]$. Also, 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.

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 neighbourhood of θ_0.

ii. $\sup_{0 \le \epsilon \le 1} \sup_{\|\theta - \theta_0\| \le \epsilon} \max_{i \ge 1} E\left[|h_m(x_{i1}'\theta)|^{4+\eta} \right] = O(m)$ as $m \to \infty$, 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 neighbourhood of θ_0, U_{θ_0}, such that

$$|g(x'\theta_1) - g(x'\theta_0)| \le M(x) \|\theta_1 - \theta_0\|,$$

where $\theta_1 \in U_{\theta_0}$. Moreover, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} (M(x_{it}))^2 = O_p(1)$.

Assumption 4.i-ii are in spirit the same as Assumption 3.i-ii. Similar to the arguments for Assumption 3.ii, we can show that Assumption 4.ii is verifiable. For example, if $x_{it}'\theta$ follows a normal distribution, then we can show that Assumption 4.ii is verifiable by going through the similar procedure of the example given for Assumption 3.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). We put Lipschitz condition on a neighbour of θ_0 rather than assume $\mathbb{X}$ is compact. In this sense, this assumption is more general compared to Ichimura (1993). The last equation in Assumption 4.iii can be easily verified under certain restriction by following the similar procedure to the second result of Lemma 2 in the appendix.

Under the extra conditions, we establish the following theorem.

Theorem 5 *Under the conditions of Theorem 2 and Assumption 4, as* $(N, T) \to (\infty, \infty)$ *jointly, then*

$$\left\|\hat{\mathcal{C}} - \mathcal{C}\right\|^2 = O_p\left(\frac{k^{3/2}}{NT}\right) + o_p\left(k^{-r}\right).$$

Similar to the discussion under Theorem 1, if we use a stronger assumption to bound the moments of $h_m(x_{it}'\theta)$ uniformly, the first term in the convergency rate above will become the optimal rate $O_p\left(\frac{k}{NT}\right)$.

Notice that the second decomposition (2.5) raises the curse of dimensionality issue when we estimate β (cf. see the convergence rate in Theorem 1), but this issue does not exist in the convergency rate given by Theorem 5.

Intuitively, we denote an estimator of $g(w)$ similar to (4.1) as

$$\hat{g}_1(w) = H(w)'\hat{\mathcal{C}} + \tilde{c}_0, \tag{4.3}$$

where $\tilde{c}_0 = -\frac{1}{NT}\sum_{i=1}^N \sum_{t=1}^T H\left(x_{it}'\hat{\theta}\right)'\hat{\mathcal{C}}$.

The integrated mean squared error of the nonparametric estimate is summarised below.

Corollary 2 *Under the conditions of Theorem 5, if in additional Assumption 2.ii holds, as* $(N, T) \to (\infty, \infty)$ *jointly, then*

$$\int_{\mathbb{R}} \left(\hat{g}_1(w) - g(w)\right)^2 \cdot \exp\left(-w^2/2\right) dw = O_p\left(\frac{k^{3/2}}{NT}\right) + o_p\left(k^{-r}\right).$$

The proofs of Theorem 5 and Corollary 2 are given in the appendix. 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 two empirical studies. In the simulation, we consider an exponential functional form, $g(w) = \exp(w)$. The expectation $E\left[\exp(x'\theta_0)\right]$ is certainly not zero, but it will not affect us to obtain a consistent estimation on θ_0. It further backs up our argument for Assumption 2.ii. Similar discussion can be found in the Monte Carlo study section of Su and Jin (2012). In empirical studies, we investigate UK's climate data and US cigarettes demand. It clearly shows that our method outperforms some existing methods in practice.

5.1 Monte Carlo Study

The data generating process (DGP) is as follows.

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

and for $j = 1, 2$

$$x_{j,it} = \rho_{x_j} x_{ji,t-1} + i.i.d.\ N(0,1) \ \text{ for } t = -99, \ldots, 0, \ldots, T,$$
$$\rho_{x_1} = 0.7, \quad \rho_{x_2} = 0.3, \quad x_{ji,-100} = 0.$$

To introduce the cross-sectional dependence to the model, we follow the DGP in Chen et al. (2012a) and let $e_t = (e_{1t}, \ldots, e_{Nt})'$, $e_{-100} = 0_{N\times 1}$ and $\rho_e = 0.2$ for $1 \le t \le T$. Then the error term e_t is generated as

$$e_t = \rho_e e_{t-1} + i.i.d.\ N\left(0_{N\times 1}, \Sigma_e\right) \ \text{ for } t = -99, \ldots, 0, \ldots, T,$$

where $\Sigma_e = (\sigma_{ij})_{N\times N} = 0.5^{|i-j|}$ for $1 \le i, j \le N$. The fixed effects, γ_i's, follow from $i.i.d.\ U(0,1)$.

The values of θ_{01} and θ_{02} are set to 0.8 and -0.6, and they are estimated by $\widehat{\theta}_1$ and $\widehat{\theta}_2$, respectively. In this Monte Carlo study, we choose $N, T = 20, 50, 100, 200$ and k as $\lfloor 1.36 \cdot \sqrt[3]{50} \rfloor = 5$, $\lfloor 1.36 \cdot \sqrt[3]{100} \rfloor = 6$ and $\lfloor 1.36 \cdot \sqrt[3]{200} \rfloor = 7$ respectively. We repeat the estimation procedure 10000 times.

As Tables 1-3 shows, all the results are very accurate. The biases and the root mean squared errors (RMSE) of $\widehat{\theta}_1$ and $\widehat{\theta}_2$ decrease as both N and T increase. Notice that the biases for this simulation are quite small, which is due to the next reasons. In (A.11), it is easy to be seen that the first term on the right hand side (RHS) is unbiased and has expectation zero. The second term on RHS of (A.11) is biased and its convergence rate is

$o_p(k^{-r})$, where r is directly related to the smoothness of the link function $g(x)$. We know that the n^{th} derivative of the exponential function exists for all positive integers $n = 1, 2, \ldots$, so it is reasonable to expect this term will generate very small bias. Similarly, we do not expect the second term on RHS of (A.11) contributes too much to RMSE.

		$\hat{\theta}_1$				$\hat{\theta}_2$			
$k = 5$	$N \setminus T$	20	50	100	200	20	50	100	200
Bias	20	-0.0012	0.0005	0.0006	0.0005	0.0000	0.0012	0.0011	0.0009
	50	-0.0004	0.0005	0.0005	0.0003	0.0001	0.0009	0.0008	0.0005
	100	-0.0005	0.0002	0.0004	0.0004	-0.0002	0.0005	0.0006	0.0004
	200	-0.0004	0.0002	0.0002	0.0001	-0.0003	0.0004	0.0003	0.0002
RMSE	20	0.0264	0.0152	0.0109	0.0083	0.0352	0.0204	0.0146	0.0111
	50	0.0171	0.0106	0.0078	0.0062	0.0228	0.0162	0.0104	0.0082
	100	0.0131	0.0081	0.0061	0.0047	0.0174	0.0108	0.0081	0.0063
	200	0.0099	0.0063	0.0048	0.0037	0.0132	0.0084	0.0064	0.0049

Table 1: Bias and RMSE for $k = 5$

		$\hat{\theta}_1$				$\hat{\theta}_2$			
$k = 6$	$N \setminus T$	20	50	100	200	20	50	100	200
Bias	20	-0.0015	-0.0002	-0.0001	-0.0002	-0.0003	0.0002	0.0001	-0.0001
	50	-0.0009	-0.0002	-0.0001	-0.0002	-0.0007	-0.0001	0.0000	-0.0002
	100	-0.0009	-0.0002	-0.0001	-0.0001	-0.0010	-0.0002	-0.0001	-0.0002
	200	-0.0009	-0.0002	-0.0001	-0.0001	-0.0010	-0.0002	-0.0002	-0.0001
RMSE	20	0.0027	0.0138	0.0093	0.0064	0.0374	0.0185	0.0123	0.0086
	50	0.0157	0.0087	0.0060	0.0042	0.0209	0.0116	0.0079	0.0055
	100	0.0109	0.0061	0.0042	0.0031	0.0145	0.0082	0.0056	0.0041
	200	0.0078	0.0045	0.0031	0.0023	0.0103	0.0060	0.0042	0.0031

Table 2: Bias and RMSE for $k = 6$

$k=7$	$N \setminus T$	$\hat{\theta}_1$ 20	50	100	200	$\hat{\theta}_2$ 20	50	100	200
Bias	20	-0.0014	-0.0001	-0.0001	-0.0001	0.0008	0.0003	0.0001	-0.0001
	50	-0.0009	-0.0002	-0.0001	-0.0002	-0.0006	-0.0001	0.0000	-0.0002
	100	-0.0009	-0.0002	-0.0001	-0.0001	-0.0010	-0.0002	-0.0001	-0.0002
	200	-0.0009	-0.0002	-0.0002	-0.0001	-0.0010	-0.0002	-0.0002	-0.0001
RMSE	20	0.0333	0.0142	0.0092	0.0063	0.0458	0.0190	0.0123	0.0084
	50	0.0161	0.0086	0.0058	0.0040	0.0214	0.0115	0.0077	0.0053
	100	0.0109	0.0060	0.0040	0.0028	0.0145	0.0079	0.0054	0.0038
	200	0.0076	0.0042	0.0029	0.0020	0.0101	0.0056	0.0039	0.0027

Table 3: Bias and RMSE for $k = 7$

5.2 Empirical Studies

In this section, we provide two empirical studies to demonstrate how our method performs in practice. As a comparison, we also run OLS regression on the following linear model after within transformation for the next two data sets.

$$y_{it} = x_{it}'\theta_0 + \gamma_i + e_{it}. \tag{5.2}$$

According to the results on convergence rates in previous sections, it is impossible to tell what the optimal k should be. To choose the truncation parameter k, we use the extended version of the generalized cross-validation (GCV) criterion proposed in Gao et al. (2002) at first. Then select a k according to the other measurements (e.g. R^2) in a small range of the $\hat{k}$ suggested by GCV. As they mention in the paper, how to select an optimum truncation parameter has not been completely solved yet.

Below SIM and LIM denote the semiparametric single-index model (2.1) and the linear model (5.2), respectively. The corresponding standard deviations are reported in the brackets.

5.2.1 UK's Climate Data

Firstly, we use the exactly same UK's climate data as Chen et al. (2013a), which is available from http://www.metoffice.gov.uk/climate/uk/stationdata/. We investigate how the average maximum temperature (TMAX) is affected by the number of millimeters of rainfall

(RAIN) and the number of hours of sunshine (SUN). The data were collected over the decade of January 1999 to December 2008 from 16 stations across UK, so $N = 16$ and $T = 120$.

The results are reported in Table 4 and Figures 1 and 2.

SIM			LIM		
$\hat{\theta}_1$	$\hat{\theta}_2$	R^2	$\hat{\theta}_1$	$\hat{\theta}_2$	R^2
0.313	0.950	0.685	0.019	0.070	0.655
(0.702)	(0.953)		(0.003)	(0.004)	

Table 4: Estimated coefficients for UK's climate data

The R^2's indicate that the semiparametric estimator proposed in this paper generates more accurate results. Compared to the $R^2 = 0.6199$ in Chen et al. (2013a), our method performs better. For our model, the number of Hermite Polynomial function is chosen as 6 (such that $h_0, h_1, \ldots, h_5$ are chosen and $K = 20$ by (2.9)). Due to the similarity, we only report the temperature plots for one station in Figure 1 and omit the others. The dash-dot line is the observed TMAX data; the solid line is the estimated temperature by our approach; and the two dash lines are 95% confidence interval obtained by using Theorem 4. Figure 1 shows that our estimates clearly capture the movement of average maximum temperature. In Figure 2, the estimated curve is plotted according to (4.3). As one can see, the linear model tells an unconvincing story. According to Figure 2 and the results from OLS, one would have concluded that as the amount of rain fall goes up, the average maximum temperature will increase. However, this seems to be very misleading. On the other hand, the single-index model tells us that the maximum temperature will decrease as the amount of rain fall increases, which is more meaningful to us.

5.2.2 US Cigarettes Demand

The data set of the second case study is from Baltagi et al. (2000) for analysing the demand for cigarettes in the U.S., who use the next linear model of the form

$$\ln C_{it} = \beta_0 + \beta_1 \ln C_{i,t-1} + \beta_2 \ln DI_{it} + \beta_3 \ln P_{it} + \beta_4 \ln PN_{it} + u_{it}, \tag{5.3}$$

where $i = 1, \ldots, 46$ and $t = 1, \ldots, 30$ represent the states and the years (1963-1992) respectively, 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

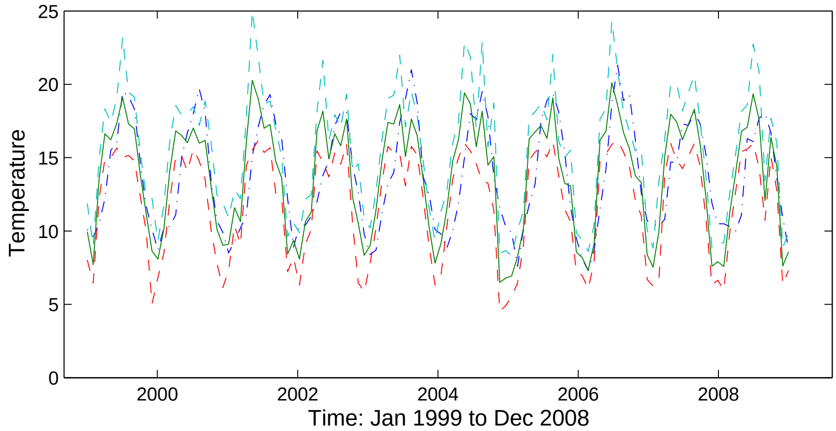

Figure 1: Estimated average maximum temperature

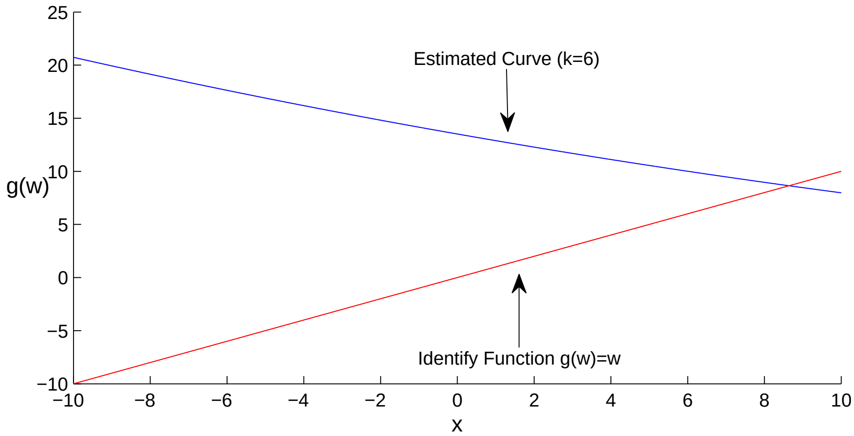

Figure 2: Estimated curve for UK's climate data

real terms, PN_{it} is the minimum real price of cigarettes in any neighbouring state and u_{it} is the disturbance term.

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

$$\ln C_{it} = g(x_{it}'\theta) + \gamma_i + e_{it}, \tag{5.4}$$

where $x_{it} = (\ln C_{i,t-1}, \ln DI_{it}, \ln P_{it}, \ln PN_{it})'$. Due to the lagged dependent value included in x_{it}, the length of time series used in the regression is 29 (such that $t = 2, \ldots, 30$). γ_i's capture all the state-specific effects. All the errors' cross-sectional dependences and year-specific effects are absorbed in e_{it}. Similar to the previous section, we report the estimates below. The results of several other attempts can be found in Baltagi et al. (2000), Mammen et al. (2009) and Chen et al. (2013b).

Compared to the $R^2 = 0.9698$ in Chen et al. (2013b), our method provides slightly better results. For our model, the number of Hermite Polynomial function is chosen as 2 (such that

	$\hat{\theta}_1$	$\hat{\theta}_2$	$\hat{\theta}_3$	$\hat{\theta}_4$	R^2
SIM	0.942	0.155	-0.288	0.070	0.973
	(0.665)	(0.412)	(0.860)	(0.776)	
LIM	0.811	0.133	-0.248	0.061	0.753
	(0.033)	(0.018)	(0.029)	(0.029)	

Table 5: Estimated coefficients for US cigarette demand

h_0 and h_1 are chosen) so that the link function g is a linear function (with a different slope compared to the identity function). Due to the similarity, we only report the plots for one state in Figure 3 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; the two dash lines are 95% confidence interval obtained by using Theorem 4. In Figure 4, the estimated curve is plotted.

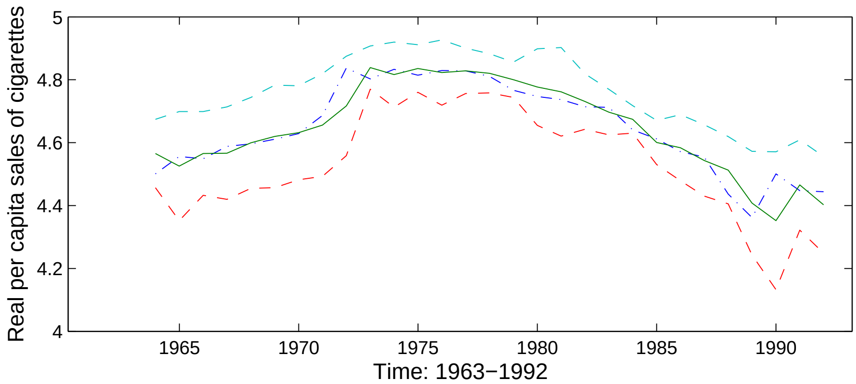

Figure 3: Estimated log real per capita sales of cigarettes

6 Conclusion

In this paper, we have proposed a semiparametric single–index panel data model associated with cross-sectional dependence, high–dimensionality, stationarity and unobservable heterogeneity. Some closed–form estimates have been proposed and the closed–form estimates have been used to recover the estimates of the parameters of interest and the link function respectively. 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 model and estimation method perform well when cross-sectional dependence

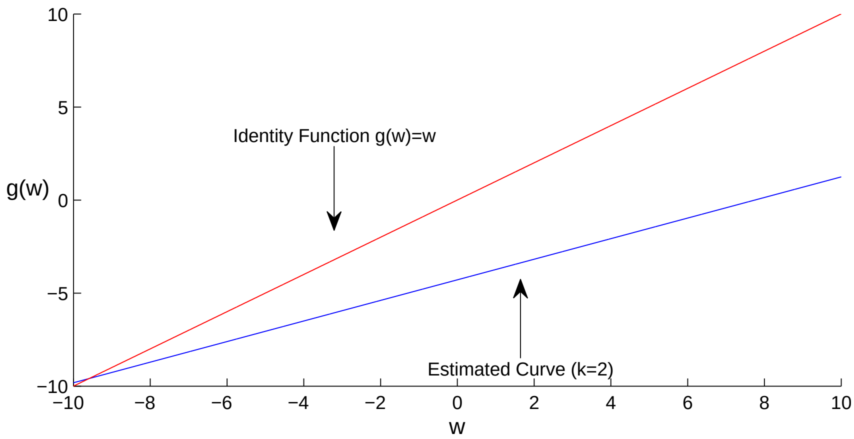

Figure 4: Estimated curve for US cigarette demand

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. Two empirical examples have shown that the proposed model and estimation method outperform some natural competitors.

Appendix

We now provide some useful lemmas before we prove the main results of this paper. Lemma 1 is in the same spirit as Lemma 12.4.2 of Blower (2009).

Lemma 1 *Suppose that* $u = (u_1, \ldots, u_d), v = (v_1, \ldots, v_d) \in \mathbb{R}^d$ *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 $p = (p_1, \ldots, p_d)$, p_j *for* $j = 1, \ldots, d$ *are all nonnegative integers,* $|p| = p_1 + \cdots + p_d$ *and* $\binom{m}{p} = \frac{m!}{\prod p_j!}$.

Proof of Lemma 1: It is known that Hermite polynomial system has the following generating function

$$\exp\left(\lambda x - \lambda^2/2\right) = \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} H_n(x). \tag{A.1}$$

For each $j = 1, \ldots, d$, by (A.1) we have $\exp\left(v_j u_j - v_j^2/2\right) = \sum_{p_j=0}^{\infty} \frac{v_j^{p_j}}{p_j!} H_{p_j}(u_j)$.

Hence, we can take product of j to obtain that

$$\exp\left(u'v - \|v\|^2/2\right) = \prod_{j=1}^{d}\sum_{p_j=0}^{\infty}\frac{v_j^{p_j}}{p_j!}H_{p_j}(u_j)$$
$$= \sum_{m=0}^{\infty}\sum_{|p|=m}\frac{1}{\prod_{j=1}^{d}p_j!}\prod_{j=1}^{d}H_{p_j}(u_j)\prod_{j=1}^{d}v_j^{p_j}.$$

Notice that $\|v\| = 1$ and once again the generating function indicates that the term of degree m on left hand side (LHS) is $\frac{1}{m!}H_m(u'v)$, which, after matching with the term of degree m on right hand side (RHS), gives the result. ■

Lemma 2 *Let Assumptions 1, 2 and 3 hold. Then, we have*

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. $\lambda_{min}\left(\frac{1}{NT}\mathcal{Z}'(I_N\otimes M_{i_T})\mathcal{Z}\right) \geq \lambda_{min}(Q_1 - Q_2)/2 > 0$.

Proof of Lemma 2: 1). Write

$$E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z(x_{it})Z(x_{it})' - Q_1\right\|^2$$
$$= \sum_{u=1}^{K}\sum_{v=1}^{K}E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_u(x_{it})z_v(x_{it}) - \frac{1}{N}\sum_{i=1}^{N}Q_{1,iuv}\right]^2, \tag{A.2}$$

where $z_u(\cdot)$ and $z_v(\cdot)$ are the u^{th} and v^{th} elements of $Z(\cdot)$, respectively, and $Q_{1,iuv}$ is the $(u,v)^{th}$ element of $Q_{1,i}$.

Observe that

$$
\begin{aligned}
& E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} z_u\left(x_{it}\right) z_v\left(x_{it}\right) - \frac{1}{N}\sum_{i=1}^{N} Q_{1,iuv}\right]^2 \\
= & \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N} E\left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\left(z_u\left(x_{it_1}\right) z_v\left(x_{it_1}\right) - Q_{1,iuv}\right)\left(z_u\left(x_{jt_2}\right) z_v\left(x_{jt_2}\right) - Q_{1,juv}\right)\right] \\
= & \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N} \mathrm{Cov}\left(z_u\left(x_{i1}\right) z_v\left(x_{i1}\right), z_u\left(x_{j1}\right) z_v\left(x_{j1}\right)\right) \\
& + \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right) \mathrm{Cov}\left(z_u\left(x_{i1}\right) z_v\left(x_{i1}\right), z_u\left(x_{j,1+t}\right) z_v\left(x_{j,1+t}\right)\right) \\
& + \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right) \mathrm{Cov}\left(z_u\left(x_{j1}\right) z_v\left(x_{j1}\right), z_u\left(x_{i,1+t}\right) z_v\left(x_{i,1+t}\right)\right) \\
= & \frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\Phi_{ijuv,1} + \Phi_{ijuv,2} + \Phi_{ijuv,3}). \qquad \text{(A.3)}
\end{aligned}
$$

We then consider each term on right hand side (RHS) of (A.3) respectively. Due to the Davydov inequality (cf. pages 19-20 in Bosq (1996) and supplementary of Su and Jin (2012))

$$
\begin{aligned}
\left|\Phi_{ijuv,2}\right| = & \left|\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right) \mathrm{Cov}\left(z_u\left(x_{i1}\right) z_v\left(x_{i1}\right), z_u\left(x_{j,1+t}\right) z_v\left(x_{j,1+t}\right)\right)\right| \\
\leq & \, c_\eta \sum_{t=1}^{T-1}\left|1-\frac{t}{T}\right| \cdot \left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)} \cdot \left(E\left[\left|z_u\left(x_{i1}\right) z_v\left(x_{i1}\right)\right|^{2+\eta/2}\right]\right)^{2/(4+\eta)} \\
& \cdot \left(E\left[\left|z_u\left(x_{j1}\right) z_v\left(x_{j1}\right)\right|^{2+\eta/2}\right]\right)^{2/(4+\eta)} \\
\leq & \, \frac{c_\eta}{2} \sum_{t=1}^{T-1}\left|1-\frac{t}{T}\right| \cdot \left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)} \cdot \left(E\left[\left|z_u\left(x_{i1}\right) z_v\left(x_{i1}\right)\right|^{2+\eta/2}\right]\right)^{4/(4+\eta)} \\
& + \frac{c_\eta}{2} \sum_{t=1}^{T-1}\left|1-\frac{t}{T}\right| \cdot \left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)} \cdot \left(E\left[\left|z_u\left(x_{j1}\right) z_v\left(x_{j1}\right)\right|^{2+\eta/2}\right]\right)^{4/(4+\eta)} \\
\leq & \, \frac{c_\eta}{2} \sum_{t=1}^{T-1} \left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)} \cdot \left(E\left[\left|z_u\left(x_{i1}\right)\right|^{4+\eta}\right] E\left[\left|z_v\left(x_{i1}\right)\right|^{4+\eta}\right]\right)^{2/(4+\eta)} \\
& + \frac{c_\eta}{2} \sum_{t=1}^{T-1} \left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)} \cdot \left(E\left[\left|z_u\left(x_{j1}\right)\right|^{4+\eta}\right] E\left[\left|z_v\left(x_{j1}\right)\right|^{4+\eta}\right]\right)^{2/(4+\eta)} \qquad \text{(A.4)}
\end{aligned}
$$

where $c_\eta = 2^{(4+2\eta)/(4+\eta)} \cdot (4+\eta)/\eta$.

In connection with Assumption 3.ii,

$$
\begin{aligned}
&\frac{1}{N^2T}\sum_{u=1}^{K}\sum_{v=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\Phi_{ijuv,2}\right| \\
\leq\; &\frac{C}{N^2T}\sum_{u=1}^{K}\sum_{v=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)}\cdot\left(O\left(\left|p_u\right|^d\right)\cdot O\left(\left|p_v\right|^d\right)\right)^{2/(4+\eta)} \\
\leq\; &\frac{C}{N^2T}\sum_{u=1}^{K}\sum_{v=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)}\cdot\left(O\left(\left|p_u\right|^d\cdot\left|p_v\right|^d\right)\right)^{1/2}=O\left(\frac{k^{3d}}{NT}\right),
\end{aligned}
$$

where the last line is true due to the fact that $\max_{1\leq u\leq K}\left|p_u\right|=k-1$ and $K=O\left(k^d\right)$.

Similarly,

$$
\frac{1}{N^2T}\sum_{u=1}^{K}\sum_{v=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\Phi_{ijuv,1}\right|=O\left(\frac{k^{3d}}{NT}\right)\text{ and }\frac{1}{N^2T}\sum_{u=1}^{K}\sum_{v=1}^{K}\sum_{i=1}^{N}\sum_{j=1}^{N}\left|\Phi_{ijuv,3}\right|=O\left(\frac{k^{3d}}{NT}\right).
$$

Thus, the result follows. ■

2). Write

$$
E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}g\left(x_{it}'\theta_0\right)\right]^2=\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}g\left(x_{it_1}'\theta_0\right)g\left(x_{jt_2}'\theta_0\right)\right]. \tag{A.5}
$$

Expanding the RHS of the above equation by the same procedure as (A.3) and (A.4), the result follows from Assumptions 1.i and 2.ii. ■

3). By following the same procedure as the first result of this lemma, the result follows. ■

4). Write

$$
\begin{aligned}
E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z\left(x_{it}\right)e_{it}\right\|^2&=\sum_{u=1}^{K}E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}z_u\left(x_{it}\right)e_{it}\right]^2 \\
&=\sum_{u=1}^{K}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}z_u\left(x_{it_1}\right)e_{it_1}z_u\left(x_{jt_2}\right)e_{jt_2}\right].
\end{aligned} \tag{A.6}
$$

Following the same procedure as the first result of this lemma, the result follows. ■

5) Write

$$
\begin{aligned}
&\lambda_{min}\left(\frac{1}{NT}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right)=\lambda_{min}\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}'\right) \\
=\;&\min_{\|\chi\|=1}\left\{\chi'\left(Q_1-Q_2\right)\chi+\chi'\left(\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{Z}_{it}\tilde{Z}_{it}'-\left(Q_1-Q_2\right)\right)\chi\right\} \\
\geq\;&\lambda_{min}\left(Q_1-Q_2\right)-\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{Z}_{it}\tilde{Z}_{it}'-\left(Q_1-Q_2\right)\right\|.
\end{aligned} \tag{A.7}
$$

We now consider $\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{Z}_{it}\tilde{Z}_{it}' - (Q_1 - Q_2)$.

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{Z}_{it}\tilde{Z}_{it}' - (Q_1 - Q_2) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Z(x_{it})Z(x_{it})' - Q_{1,i}\right) - \frac{1}{N}\sum_{i=1}^{N}\left(\bar{Z}_{i.}\bar{Z}_{i.}' - q_i q_i'\right)$$

Similar to the first result of this lemma

$$\begin{aligned}&\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{Z}_{i.}\bar{Z}_{i.}' - q_i q_i'\right)\right\| \\ \leq &\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{Z}_{i.} - q_i\right)\left(\bar{Z}_{i.} - q_i\right)'\right\| + \left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{Z}_{i.} - q_i\right)q_i'\right\| + \left\|\frac{1}{N}\sum_{i=1}^{N}q_i\left(\bar{Z}_{i.} - q_i\right)'\right\| = o_p(1).\end{aligned}$$

In connection with the first result of this lemma, we obtain

$$\begin{aligned}&\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{Z}_{it}\tilde{Z}_{it}' - (Q_1 - Q_2)\right\| \\ \leq &\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(Z(x_{it})Z(x_{it})' - Q_{1,i}\right)\right\| + \left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{Z}_{i.}\bar{Z}_{i.}' - q_i q_i'\right)\right\| = o_p(1).\end{aligned}$$

Thus, the result follows. ■

Lemma 3 *Let Assumptions 1–4 hold. Then the following results hold uniformly in a small neighbour of θ_0:*

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. $\lambda_{min}\left(\frac{1}{NT}\mathcal{H}(\theta)'(I_N \otimes M_{i_T})\mathcal{H}(\theta)\right) \geq \lambda_{min}(R_1(\theta) - R_2(\theta))/2 > 0$.

Proof of Lemma 3: 1). Write

$$\begin{aligned}&E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}H\left(x_{it}'\theta\right)H\left(x_{it}'\theta\right)' - R_1(\theta)\right\|^2 \\ =&\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}h_u\left(x_{it}'\theta\right)h_v\left(x_{it}'\theta\right) - \frac{1}{N}\sum_{i=1}^{N}R_{1,iuv}(\theta)\right]^2,\end{aligned} \tag{A.8}$$

where $h_u(\cdot)$ and $h_v(\cdot)$ are the u^{th} and v^{th} elements of $H(\cdot)$, respectively, and $R_{1,iuv}$ is the $(u,v)^{th}$ element of $R_{1,i}(\theta)$.

Observe that

$$E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}h_u\left(x_{it}'\theta\right)h_v\left(x_{it}'\theta\right)-\frac{1}{N}\sum_{i=1}^{N}R_{1,iuv}\left(\theta\right)\right]^2$$

$$=\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\mathrm{Cov}\left(h_u\left(x_{i1}'\theta\right)h_v\left(x_{i1}'\theta\right),h_u\left(x_{j1}'\theta\right)h_v\left(x_{j1}'\theta\right)\right)$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\mathrm{Cov}\left(h_u\left(x_{i1}'\theta\right)h_v\left(x_{i1}'\theta\right),h_u\left(x_{j,1+t}'\theta\right)h_v\left(x_{j,1+t}'\theta\right)\right)$$

$$+\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\mathrm{Cov}\left(h_u\left(x_{j1}'\theta\right)h_v\left(x_{j1}'\theta\right),h_u\left(x_{i,1+t}'\theta\right)h_v\left(x_{i,1+t}'\theta\right)\right)$$

$$=\frac{1}{N^2T}\sum_{i=1}^{N}\sum_{j=1}^{N}(\Psi_{ijuv,1}(\theta)+\Psi_{ijuv,2}(\theta)+\Psi_{ijuv,3}(\theta)). \tag{A.9}$$

By the similar procedure of (A.4)

$$|\Psi_{ijuv,2}(\theta)|=\left|\sum_{t=1}^{T-1}\left(1-\frac{t}{T}\right)\mathrm{Cov}\left(h_u\left(x_{i1}'\theta\right)h_v\left(x_{i1}'\theta\right),h_u\left(x_{j,1+t}'\theta\right)h_v\left(x_{j,1+t}'\theta\right)\right)\right|$$

$$\leq\frac{c_\eta}{2}\sum_{t=1}^{T-1}\left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)}\cdot\left(E\left[\left|h_u\left(x_{i1}'\theta\right)\right|^{4+\eta}\right]E\left[\left|h_v\left(x_{i1}'\theta\right)\right|^{4+\eta}\right]\right)^{2/(4+\eta)}$$

$$+\frac{c_\eta}{2}\sum_{t=1}^{T-1}\left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)}\cdot\left(E\left[\left|h_u\left(x_{j1}'\theta\right)\right|^{4+\eta}\right]E\left[\left|h_v\left(x_{j1}'\theta\right)\right|^{4+\eta}\right]\right)^{2/(4+\eta)},$$

where $c_\eta=2^{(4+2\eta)/(4+\eta)}\cdot(4+\eta)/\eta$.

In connection with Assumption 4.ii,

$$\frac{1}{N^2T}\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\Psi_{ijuv,2}|\leq\frac{C}{N^2T}\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)}\cdot\left(O\left(u\right)\cdot O\left(v\right)\right)^{2/(4+\eta)}$$

$$\leq\frac{C}{N^2T}\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{t=1}^{T-1}\left(\alpha_{ij}\left(t\right)\right)^{\eta/(4+\eta)}\cdot\left(O\left(u\cdot v\right)\right)^{1/2}=O\left(\frac{k^3}{NT}\right).$$

Similarly,

$$\frac{1}{N^2T}\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\Psi_{ijuv,1}|=O\left(\frac{k^3}{NT}\right)\text{ and }\frac{1}{N^2T}\sum_{u=1}^{k-1}\sum_{v=1}^{k-1}\sum_{i=1}^{N}\sum_{j=1}^{N}|\Psi_{ijuv,3}|=O\left(\frac{k^3}{NT}\right).$$

Thus, the result follows. ■

2). Using the similar procedure to the first result of this lemma, the result follows. ■

3). Write

$$E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}H\left(x_{it}'\theta\right)e_{it}\right\|^2=\sum_{u=1}^{k-1}E\left[\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}h_u\left(x_{it}'\theta\right)e_{it}\right]^2$$

$$=\sum_{u=1}^{k-1}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}h_u\left(x_{it_1}'\theta\right)h_u\left(x_{jt_2}'\theta\right)e_{it_1}e_{jt_2}\right]. \tag{A.10}$$

Similar to the procedure used in proving the first result of this lemma, the result follows. ■

4) Similar to (A.7), write

$$\lambda_{min}\left(\frac{1}{NT}\mathcal{H}'\left(I_N\otimes M_{i_T}\right)\mathcal{H}\right)$$
$$\geq \lambda_{min}\left(R_1\left(\theta\right)-R_2\left(\theta\right)\right)-\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\tilde{H}\left(x_{it}'\theta\right)\tilde{H}\left(x_{it}'\theta\right)'-\left(R_1\left(\theta\right)-R_2\left(\theta\right)\right)\right\|,$$

where $\tilde{H}\left(x_{it}'\theta\right)=H\left(x_{it}'\theta\right)-\bar{H}_{i.}\left(\theta\right)$ and $\bar{H}_{i.}\left(\theta\right)=\frac{1}{T}\sum_{t=1}^{T}H\left(x_{it}'\theta\right)$.

We now consider

$$\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{H}\left(x_{it}'\theta\right)\tilde{H}\left(x_{it}'\theta\right)'-\left(R_1\left(\theta\right)-R_2\left(\theta\right)\right)$$
$$=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(H\left(x_{it}'\theta\right)H\left(x_{it}'\theta\right)'-R_{1,i}\left(\theta\right)\right)-\frac{1}{N}\sum_{i=1}^{N}\left(\bar{H}_{i.}\left(\theta\right)\bar{H}_{i.}\left(\theta\right)'-r_i\left(\theta\right)r_i\left(\theta\right)'\right).$$

In a similar fashion to the proof of the first result of this lemma, we have

$$\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{H}_{i.}\left(\theta\right)\bar{H}_{i.}\left(\theta\right)'-r_i\left(\theta\right)r_i\left(\theta\right)'\right)\right\|$$
$$\leq\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{H}_{i.}\left(\theta\right)-r_i\left(\theta\right)\right)\left(\bar{H}_{i.}\left(\theta\right)-r_i\left(\theta\right)\right)'\right\|+\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{H}_{i.}\left(\theta\right)-r_i\left(\theta\right)\right)r_i\left(\theta\right)'\right\|$$
$$+\left\|\frac{1}{N}\sum_{i=1}^{N}r_i\left(\theta\right)\left(\bar{H}_{i.}\left(\theta\right)-r_i\left(\theta\right)\right)'\right\|=o_p\left(1\right).$$

In connection with the first result of this lemma, we obtain

$$\left\|\frac{1}{NT}\sum_{t=1}^{T}\sum_{i=1}^{N}\tilde{H}\left(x_{it}'\theta\right)\tilde{H}\left(x_{it}'\theta\right)'-\left(R_1\left(\theta\right)-R_2\left(\theta\right)\right)\right\|$$
$$\leq\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}H\left(x_{it}'\theta\right)H\left(x_{it}'\theta\right)'-R_1\left(\theta\right)\right\|+\left\|\frac{1}{N}\sum_{i=1}^{N}\left(\bar{H}_{i.}\left(\theta\right)\bar{H}_{i.}\left(\theta\right)'-r_i\left(\theta\right)r_i\left(\theta\right)'\right)\right\|=o_p\left(1\right).$$

Thus, the result follows. ■

Proof of Theorem 1: We now start the proof of the consistency. By the uniqueness of the Moore-Penrose inverse and the fifth result of Lemma 2 of this appendix, the $K\times K$ dimensions matrix $\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}$ is the inverse of $\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}$ for each K. Therefore,

$$\hat{\beta}-\beta=\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{E}$$
$$+\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{D}(\theta_0). \qquad \text{(A.11)}$$

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

$$E\left\|\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}\right\|^2 = E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z\left(x_{it}\right)e_{it} - \frac{1}{N}\sum_{i=1}^{N}\bar{Z}_{i.}\bar{e}_{i.}\right\|^2$$
$$\leq 2E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z\left(x_{it}\right)e_{it}\right\|^2 + 2E\left\|\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_{i.}\bar{e}_{i.}\right\|^2. \tag{A.12}$$

By the fourth result of Lemma 2, we have $E\left\|\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z\left(x_{it}\right)e_{it}\right\|^2 = O\left(\frac{k^{3d/2}}{NT}\right)$. For the second term on RHS of (A.12), write

$$E\left\|\frac{1}{N}\sum_{i=1}^{N}\bar{Z}_{i.}\bar{e}_{i.}\right\|^2 = \sum_{u=1}^{K}E\left[\frac{1}{NT^2}\sum_{i=1}^{N}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}z_u\left(x_{it_1}\right)e_{it_2}\right]^2$$
$$= \sum_{u=1}^{K}\frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\frac{1}{T^4}\sum_{t_1=1}^{T}\sum_{t_2=1}^{T}\sum_{t_3=1}^{T}\sum_{t_4=1}^{T}z_u\left(x_{it_1}\right)e_{it_2}z_u\left(x_{jt_3}\right)e_{jt_4}\right]$$
$$= \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[\frac{1}{T^2}\sum_{t_2=1}^{T}\sum_{t_4=1}^{T}e_{it_2}e_{jt_4}\right]\cdot\sum_{u=1}^{K}E\left[\frac{1}{T^2}\sum_{t_1=1}^{T}\sum_{t_3=1}^{T}z_u\left(x_{it_1}\right)z_u\left(x_{jt_3}\right)\right],$$

where the last line follows Assumption 1.ii.

By Cauchy-Schwarz inequality, moment monotonicity and Assumption 3.ii respectively,

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

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

$$\left\|\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}\right\|^2 = O_p\left(\frac{k^{3d/2}}{NT}\right). \tag{A.13}$$

According to the fifth result of Lemma 2 and (A.13), we obtain

$$\left\|\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}\right\|^2$$
$$= \mathcal{E}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}$$
$$\leq \left[\lambda_{min}\left(\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right)\right]^{-2}\cdot\left\|\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}\right\|^2 = O_p\left(\frac{k^{3d/2}}{NT}\right). \tag{A.14}$$

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

$$\begin{aligned}
&\left\|\left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right)^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\theta_0)\right\|^2 \\
&= \mathcal{D}(\theta_0)'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}/(NT)\right)^{-} \\
&\quad \cdot \left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right)^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\theta_0)/(NT) \\
&\leq \left[\lambda_{min}\left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}/(NT)\right)\right]^{-1} \\
&\quad \cdot \mathcal{D}(\theta_0)'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right)^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\theta_0)/(NT) \\
&\leq \left[\lambda_{min}\left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}/(NT)\right)\right]^{-1} \cdot \lambda_{max}(W) \cdot \left(\|\mathcal{D}(\theta_0)\|^2/(NT)\right). \qquad \text{(A.15)}
\end{aligned}$$

Note that $W = \left(I_N \otimes M_{i_T}\right)\mathcal{Z}\left(\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right)^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)$ is symmetric and idempotent, so $\lambda_{max}(W) = 1$. According to Assumption 3.ii and the Weak Law of Large Numbers (WLLN), it is easy to know that $\|\mathcal{D}(\theta_0)\|^2/(NT) = o_p\left(k^{-r}\right)$. In connection with the fifth result of Lemma 2 of this appendix, we obtain that

$$\left\|\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\theta_0)\right\|^2 = o_p\left(k^{-r}\right). \qquad \text{(A.16)}$$

Therefore, the theorem follows form (A.14) and (A.16). ■

Proof of Theorem 2: It is easy to know that $sgn(\hat{\beta}_1) \cdot \left(\sum_{i=1}^{d} \hat{\beta}_i^2\right)^{-1/2}$ converges to $|c_1|^{-1}$ by (2.7) and Theorem 1, so we only need to consider $\sqrt{NT} \cdot Q_3\left(\hat{\beta} - \beta\right)$ and write

$$\begin{aligned}
\sqrt{NT} \cdot Q_3\left(\hat{\beta} - \beta\right) &= \sqrt{NT} \cdot Q_3\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\theta_0) \\
&\quad + \sqrt{NT} \cdot Q_3\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}. \qquad \text{(A.17)}
\end{aligned}$$

Notice that $K = O\left(k^d\right)$ and $Q_3 = O(1)$. In connection with (A.16) and the assumption in the body of this theorem, it is straightforward to obtain

$$\begin{aligned}
&\left\|\sqrt{NT} \cdot Q_3\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\theta_0)\right\| \\
&\leq \sqrt{NT} \cdot O(1) \cdot o_p\left(k^{-r/2}\right) = o_p(1). \qquad \text{(A.18)}
\end{aligned}$$

Then, to achieve the normality, we need only to consider the second term on RHS of (A.17).

$$\begin{aligned}
&\sqrt{NT} \cdot Q_3\left[\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E} \\
&= \sqrt{NT} \cdot Q_3\left(\left(\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{Z}\right)^{-} - \left(Q_1 - Q_2\right)^{-}\right)\left(\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}\right) \\
&\quad + \sqrt{NT} \cdot Q_3\left(Q_1 - Q_2\right)^{-}\left(\frac{1}{NT}\mathcal{Z}'\left(I_N \otimes M_{i_T}\right)\mathcal{E}\right) \qquad \text{(A.19)}
\end{aligned}$$

For two non-singular symmetric matrices A, B with same dimensions, we observe that by Theorem 2 on page 35 of Magnus (2007)

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

where $\lambda_{min}(A \otimes B) = \lambda_{min}(A) \cdot \lambda_{min}(B)$ by Theorem 1 on page 28 of Magnus (2007). Therefore, in connection with the proof of the fifth result of Lemma 2 in this appendix,

$$\left\| \left(\frac{1}{NT} \mathcal{Z}' \left(I_N \otimes M_{i_T} \right) \mathcal{Z} \right)^- - \left(Q_1 - Q_2 \right)^- \right\| = O_p \left(\sqrt{\frac{k^{3d}}{NT}} \right).$$

Moreover, by (A.13), we can obtain that

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

The second term on RHS of (A.19) can be written as follows.

$$\begin{aligned} & \sqrt{NT} \cdot Q_3 \left(Q_1 - Q_2 \right)^- \left(\frac{1}{NT} \mathcal{Z}' \left(I_N \otimes M_{i_T} \right) \mathcal{E} \right) \\ = & \sqrt{NT} \cdot Q_3 \left(Q_1 - Q_2 \right)^- \frac{1}{N} \sum_{i=1}^{N} \left(q_i - \bar{Z}_{i.} \right) \bar{e}_{i.} \\ & + \sqrt{NT} \cdot Q_3 \left(Q_1 - Q_2 \right)^- \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Z \left(x_{it} \right) - q_i \right) e_{it} \end{aligned} \qquad \text{(A.20)}$$

For the first term on RHS of (A.20), we have $E \left\| \frac{1}{N} \sum_{i=1}^{N} \left(q_i - \bar{Z}_{i.} \right) \bar{e}_{i.} \right\|^2 = O \left(\frac{k^{3d/2}}{NT^2} \right)$. Similar to (A.14), $\left\| \left(Q_1 - Q_2 \right)^- \frac{1}{N} \sum_{i=1}^{N} \left(q_i - \bar{Z}_{i.} \right) \bar{e}_{i.} \right\| = O_p \left(\sqrt{\frac{k^{3d/2}}{NT^2}} \right)$.

Therefore,

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

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 (A.20). Write

$$\sqrt{NT} \cdot Q_3 \left(Q_1 - Q_2 \right)^- \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(Z \left(x_{it} \right) - q_i \right) e_{it} = \sum_{t=1}^{T} V_{NT}(t), \qquad \text{(A.21)}$$

where $V_{NT}(t) = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} Q_3 \left(Q_1 - Q_2 \right)^- \left(Z \left(x_{it} \right) - q_i \right) e_{it}$.

Notice that Q_3 is just a selection 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.

Partition the set $\{1, \ldots, T\}$ 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

$$l_T = \lfloor T^{(\lambda-1)/\lambda} \rfloor, \quad s_T = \lfloor T^{1/\lambda} \rfloor, \quad \kappa_T = \lfloor T / (l_T + s_T) \rfloor \quad \text{for any } \lambda > 2.$$

For $\rho = 1, \ldots, \kappa_T$, let $\hat{V} = \sum_{t=\kappa_T(l_T+s_T)+1}^{T} V_{NT}(t)$.

$$\tilde{V}_\rho = \sum_{t=(\rho-1)(l_T+s_T)+1}^{\rho l_T+(\rho-1)s_T} V_{NT}(t) \quad \text{and} \quad \bar{V}_\rho = \sum_{t=\rho l_T+(\rho-1)s_T+1}^{\rho(l_T+s_T)} V_{NT}(t)\,.$$

For the small blocks, it can be seen

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

where $\bar{V}_\rho = \sum_{t=\rho l_T+(\rho-1)s_T+1}^{\rho(l_T+s_T)} V_{NT}(t) = (\bar{v}_{\rho,1}, \ldots, \bar{v}_{\rho,d})'$.

By the properties of α-mixing time series and a procedure similar to (A.6) in Chen et al. (2012b), we obtain

$$E\left[\sum_{\rho=1}^{\kappa_T}\left\|\bar{V}_\rho\right\|^2\right] = O\left(\frac{\kappa_T s_T}{T}\right) = o(1).$$

Analogously, we have

$$E\left\|\hat{V}\right\|^2 = O\left(\frac{T-\kappa_T l_T}{T}\right) = o(1).$$

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

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

for some $0 < C < \infty$, which implies that $\tilde{V}_\rho$ for $\rho = 1, \ldots, \kappa_T$ are asymptotically independent. Furthermore, as in the proof of Theorem 2.21.(ii) in Fan and Yao (2003), we have

$$\text{Cov}\left[\tilde{V}_1\right] = \frac{l_T}{T}\Xi_0\left(I + o(1)\right),$$

where

$$\begin{aligned}\Xi_0 = \lim_{N\to\infty}\frac{1}{N}\sum_{i=1}^{N}Q_3\left(Q_1-Q_2\right)^-&\Big\{E\left[e_{i1}^2\left(Z(x_{i1})-q_i\right)\left(Z(x_{i1})-q_i\right)'\right]\\ &+\sum_{t=2}^{\infty}E\left[e_{i1}e_{it}\left(Z(x_{i1})-q_i\right)\left(Z(x_{it})-q_i\right)'\right]\\ &+\sum_{t=2}^{\infty}E\left[e_{i1}e_{it}\left(Z(x_{it})-q_i\right)\left(Z(x_{i1})-q_i\right)'\right]\Big\}\left(Q_1'-Q_2'\right)^-Q_3'.\end{aligned}$$

It further implies that

$$\sum_{\rho=1}^{\kappa_T}\text{Cov}\left[\tilde{V}_\rho\right] = \kappa_T\cdot\text{Cov}\left[\tilde{V}_1\right] = \frac{\kappa_T l_T}{T}\Xi_0\left(I+o(1)\right) \to \Xi_0.$$

Thus, the Feller condition is satisfied.

Moreover, by Cauchy-Schwarz inequality, we have

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

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

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

By the assumption in the body of the theorem

$$E\left[\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N} Q_3\left(Q_1 - Q_2\right)^{-1}\left(Z\left(x_{i1}\right) - q_i\right)e_{i1}\right\|^4\right] = O(1).$$

Therefore, $E\left[\left\|\tilde{V}_\rho\right\|^3\right] = O\left(\left(\frac{l_T}{T}\right)^{3/2}\right)$, which implies that

$$E\left[\left\|\tilde{V}_\rho\right\|^2 \cdot I\left\{\|V_\rho\| \geq \varepsilon\right\}\right] \leq O\left(\left(\frac{l_T}{T}\right)^{4/3}\right) = o\left(\frac{l_T}{T}\right).$$

Consequently,

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

Therefore, the Lindeberg condition is satisfied. Therefore, the proof is completed. ■

Proof of Theorem 3: By (2.5), we have the following decomposition:

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

Moreover, $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} g\left(x_{it}'\theta_0\right) = O_p\left(\frac{1}{\sqrt{NT}}\right)$ by the second result of Lemma 2. Plus $\hat{c}_0$ from both sides and organize the equation, so we obtain that

$$\hat{c}_0 - c_0 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)'\left(\beta - \hat{\beta}\right) + \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_k\left(x_{it}'\theta_0\right) + O_p\left(\frac{1}{\sqrt{NT}}\right). \quad \text{(A.22)}$$

In view of the fact that $\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)'\right)$ has rank one and using the similar procedure of (A.14), it may be shown

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)'\left(\beta - \hat{\beta}\right)\right)^2$$
$$= \left(\hat{\beta} - \beta\right)'\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} Z\left(x_{it}\right)'\right)\left(\hat{\beta} - \beta\right)$$
$$\leq C \cdot \left\|\hat{\beta} - \beta\right\|^2 = O_p\left(\frac{k^{3d/2}}{NT}\right) + o_p\left(k^{-r}\right).$$

By using Cauchy-Schwarz inequality twice

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\delta_k\left(x_{it}'\theta_0\right)\right)^2 \leq \left[\sum_{i=1}^{N}\left(\frac{1}{\sqrt{NT}}\sum_{t=1}^{T}\delta_k\left(x_{it}'\theta_0\right)\right)^2\right]\cdot\left[\sum_{i=1}^{N}\left(\frac{1}{\sqrt{N}}\right)^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\frac{1}{T}\sum_{t=1}^{T}\delta_k\left(x_{it}'\theta_0\right)\right)^2 \leq \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\delta_k\left(x_{it}'\theta_0\right)\right)^2.$$

Moreover, we have shown that $\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\delta_k\left(x_{it}'\theta_0\right)\right)^2 = o_p\left(k^{-r}\right)$ in the proof of Theorem 1.

Based on the above, the result has been proved. ■

Proof of Theorem 4: By (A.22) and the assumptions in the body of this theorem, it is easy to obtain the next equation after some algebra.

$$\begin{aligned}
&\sqrt{\frac{NT}{K^{3/2}}}\left(\hat{g}\left(x'\hat{\theta}\right)-g(x'\theta_0)\right)\\
&=\sqrt{\frac{NT}{K^{3/2}}}Z_{NT}(x)'\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{D}(\theta_0)\\
&+\sqrt{\frac{NT}{K^{3/2}}}Z_{NT}(x)'\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{E}+o_p(1),
\end{aligned} \tag{A.23}$$

where $Z_{NT}(x) = \left(Z\left(x\right) - \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}Z\left(x_{it}\right)\right)$.

In connection with (A.16), it is straightforward to obtain that

$$\begin{aligned}
&\left\|\sqrt{\frac{NT}{K^{3/2}}}Z_{NT}(x)'\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{D}(\theta_0)\right\|\\
&\leq\sqrt{\frac{NT}{K^{3/2}}}\cdot O_p(\sqrt{k^{3d/2}})\cdot O_p\left(k^{-r/2}\right) = O_p\left(\sqrt{\frac{NT}{k^r}}\right) = o_p(1).
\end{aligned} \tag{A.24}$$

Thus, to prove the normality, we need only to consider the second term on RHS of (A.23):

$$\begin{aligned}
&\sqrt{\frac{NT}{K^{3/2}}}Z_{NT}(x)'\left[\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right]^{-}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{E}\\
&=\sqrt{\frac{NT}{K^{3/2}}}Z_{NT}(x)'\left(\left(\frac{1}{NT}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{Z}\right)^{-}-\left(Q_1-Q_2\right)^{-}\right)\left(\frac{1}{NT}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{E}\right)\\
&+\sqrt{\frac{NT}{K^{3/2}}}Z_{NT}(x)'\left(Q_1-Q_2\right)^{-}\left(\frac{1}{NT}\mathcal{Z}'\left(I_N\otimes M_{i_T}\right)\mathcal{E}\right).
\end{aligned} \tag{A.25}$$

Similar to the proof procedure of Theorem 2, write

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

Similarly, we can show

$$\begin{aligned}
&\sqrt{\frac{NT}{K^{3/2}}} Z_{NT}(x)' (Q_1 - Q_2)^- \left(\frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{E} \right) \\
&= \sqrt{\frac{NT}{K^{3/2}}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \left(\frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{E} \right) + O_p \left(\sqrt{\frac{k^{3d/2}}{T}} \right).
\end{aligned}$$

Thus, we just need to focus on the next term:

$$\begin{aligned}
&\sqrt{\frac{NT}{K^{3/2}}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \left(\frac{1}{NT} \mathcal{Z}' (I_N \otimes M_{i_T}) \mathcal{E} \right) \\
&= \sqrt{\frac{NT}{K^{3/2}}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \frac{1}{N} \sum_{i=1}^{N} \left(q_i - \bar{Z}_{i.} \right) \bar{e}_{i.} \\
&+ \sqrt{\frac{NT}{K^{3/2}}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (Z(x_{it}) - q_i) e_{it}. \qquad \text{(A.26)}
\end{aligned}$$

In Theorem 2, we have shown that

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

Hence, we obtain

$$\begin{aligned}
&\left\| \sqrt{\frac{NT}{K^{3/2}}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \frac{1}{N} \sum_{i=1}^{N} \left(q_i - \bar{Z}_{i.} \right) \bar{e}_{i.} \right\| \\
&\leq \sqrt{\frac{NT}{K^{3/2}}} \cdot O \left(\sqrt{k^{3d/2}} \right) \cdot O_p \left(\sqrt{\frac{k^{3d/2}}{NT^2}} \right) = o_p(1).
\end{aligned}$$

We still use the large-block and small-block technique to prove the normality for the second term on RHS of (A.26). Write

$$\sqrt{\frac{NT}{K^{3/2}}} (Z(x) - \bar{q})' (Q_1 - Q_2)^- \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} (Z(x_{it}) - q_i) e_{it} = \sum_{t=1}^{T} \mathcal{V}_{NTK}(t), \qquad \text{(A.27)}$$

where

$$\mathcal{V}_{NTK}(t) = \frac{1}{\sqrt{NTK^{3/2}}} \sum_{i=1}^{N} (Z(x) - \bar{q})' (Q_1 - Q_2)^- (Z(x_{it}) - q_i) e_{it}.$$

Notice that

$$\begin{aligned}
&\left((Z(x) - \bar{q})' (Q_1 - Q_2)^- (Z(x_{it}) - q_i) e_{it} \right)^2 \\
&\leq \lambda_{max} \left((Z(x) - \bar{q})(Z(x) - \bar{q})' \right) \cdot \left\| (Q_1 - Q_2)^- (Z(x_{it}) - q_i) e_{it} \right\|^2 \\
&= O_p(k^{3d/2}),
\end{aligned}$$

so that $\mathcal{V}_{NTK}$ is a summation of random scalar and its absolute value is bounded uniformly in K with probability one. Then the conventional CLT is applicable. The rest of the proof will be exactly the same as that of Theorem 2 of this paper and that of Lemma A.1 in Chen et al. (2012b), so we omit them there. ■

Proof of Theorem 5: By the uniqueness of the Moore-Penrose inverse and the fourth result of Lemma 3 of this appendix above, the $(k-1)\times(k-1)$ dimensions matrix $\left[\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\right]^{-}$ is the inverse of $\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})$ for each k. Therefore,

$$\begin{aligned}
\hat{\mathcal{C}} - \mathcal{C} &= \left[\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\right]^{-} \mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{Y} \\
&\quad - \left[\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\right]^{-} \mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\mathcal{C} \\
&= \left[\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\right]^{-} \mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\left(\mathcal{G}(\theta_0) - \mathcal{G}(\hat{\theta})\right) \\
&\quad + \left[\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\right]^{-} \mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{D}(\hat{\theta}) \\
&\quad + \left[\mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{H}(\hat{\theta})\right]^{-} \mathcal{H}(\hat{\theta})'\left(I_N \otimes M_{i_T}\right)\mathcal{E},
\end{aligned} \tag{A.28}$$

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

Similar to (A.16), we have

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

By the third and fourth results of Lemma 3 and the similar procedure of (A.14), we obtain

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

Then, we need only to consider the next term. By the same proof as (A.15) and Assumption 4.iii, we write

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

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

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

Since $\tilde{W}$ is symmetric and idempotent, $\lambda_{max}(\tilde{W}) = 1$.

By Assumption 4.iii and Theorem 2, we know that

$$\frac{1}{NT}\|\mathcal{X}\|^2 \cdot \left\|\theta_0 - \hat{\theta}\right\|^2 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}(M(x_{it}))^2 \cdot \left\|\theta_0 - \hat{\theta}\right\|^2 = O_p\left(\frac{1}{NT}\right).$$

Hence, similar with (A.16), we obtain that

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

Based on the above, the result has been proved. ■

Proof of Corollary 2: Write

$$\begin{aligned}
&\int \left(\hat{g}_1(w) - g(w) \right)^2 \cdot \exp\left(-w^2/2 \right) dw \\
&= \int \left(H(w)\hat{\mathcal{C}} + \tilde{c}_0 - c_0 - H(w)\mathcal{C} - \delta_k(w) \right)^2 \cdot \exp\left(-w^2/2 \right) dw \\
&\leq 4 \int \left(\hat{\mathcal{C}} - \mathcal{C} \right)' H(w)H(w)' \left(\hat{\mathcal{C}} - \mathcal{C} \right) \cdot \exp\left(-w^2/2 \right) dw \\
&\quad +4 \left\| \tilde{c}_0 - c_0 \right\|^2 + 2 \int \delta_k(w)^2 \cdot \exp\left(-w^2/2 \right) dw \\
&= 4 \left\| \hat{\mathcal{C}} - \mathcal{C} \right\|^2 + 4 \left\| \tilde{c}_0 - c_0 \right\|^2 + 2 \int \delta_k(w)^2 \cdot \exp\left(-w^2/2 \right) dw.
\end{aligned}$$

By going through the exactly same procedure as Theorem 3, it is easy to prove that

$$\left\| \tilde{c}_0 - c_0 \right\|^2 = O_p \left(\frac{k^{3/2}}{NT} \right) + o_p \left(k^{-r} \right).$$

For the truncated residual term, it is easy to verify the standard multivariate normal density is covered by Assumption 3.ii. Therefore, $\int \delta_k(w)^2 \cdot \exp\left(-w^2/2 \right) dw = o(k^{-r})$ by using the substitution rule of integration and Assumption 3.ii. ■

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.

D. Bosq. *Nonparametric Statistics for Stochastic Processes: Estimation and Prediction*. Springer, New York, 1996.

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.

A. DasGupta. *Asymptotic Theory of Statistics and Probability.* Springer Science+Business Media, LLC, 2008.

J. Fan and Q. Yao. *Nonlinear Time Series: Nonparametric and Parametric Methods.* Springer-Verlag, 2003.

J. Gao. *Nonlinear Time Series: Semiparametric and Nonparametric Methods.* Chapman and Hall, 2007.

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. *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley & Sons Ltd, third edition, 2007.

E. Mammen, B. Støve, and D. Tjøstheim. Nonparametric additive models for panels of time series. *Econometric Theory*, pages 442–481, 2009.

P. Nevai. Géza fredu, orthogonal polynomials and christoffel functions. A case study. *Journal of Approximation Theory*, pages 3–167, 1986.

W. K. Newey. Covergece 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.